

Tilburg University

Beheersingsleren

Weeda, Wouter Cornelis

Publication date:
1982

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Weeda, W. C. (1982). *Beheersingsleren: het model getoetst in de tijd*. [, Tilburg University]. [s.n.].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Beheersingsleren:

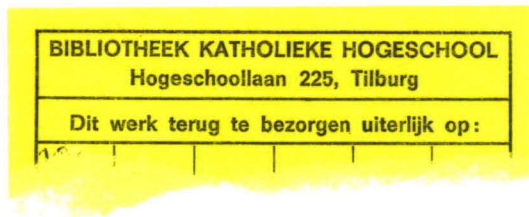
het model getoetst in de tijd.



W.C. Weeda

Beheersingsleren:

het model getoetst in de tijd.



Promotor: Prof. dr. L.F.W. de Klerk

BEHEERSINGSLEREN:
het model getoetst in de tijd

Proefschrift
ter verkrijging van de graad van
doctor in de sociale wetenschappen
aan de Katholieke Hogeschool Tilburg,
op gezag van
de rector magnificus, prof. dr. G. M. van Veldhoven,
in het openbaar te verdedigen
ten overstaan van
een door het college van decanen aangewezen commissie
in de aula van de Hogeschool
op vrijdag 29 oktober 1982
te 16.15 uur

door

WOUTER CORNELIS WEEDA

geboren te Rotterdam

412 E 9

Katholieke Hogeschool Tilburg	S. Nr. 827002
	Sig. 412 E 9
	UDC 159.953.5: 370.131.1:043.3

Aan Winny en Justin

WOORD VOORAF

Het onderzoek waarop dit proefschrift is gebaseerd, is uitgevoerd toen ik als wetenschappelijk medewerker verbonden was aan de Vakgroep Onderwijspsychologie der Katholieke Hogeschool Tilburg. Medio 1977 werd de data-verzamelingsfase afgesloten. Dat was ook het moment waarop ik een nieuwe functie aanvaardde bij het Centraal Instituut voor Toetsontwikkeling (Cito) te Arnhem als Hoofd van de afdeling Onderzoek en Psychometrische Dienstverlening (OPD). Het inwerken in een nieuw werkdomein, alsmede het vormgeven van een nieuwe afdeling heeft meer geïnterfereerd met de afronding van de verslaggeving van het onderzoek dan ik had verwacht en gehoopt.

De laatste fase waarin dit proefschrift uiteindelijk werd geschreven, heb ik als een, weliswaar veel tijd en energie vergende, maar toch zeer plezierige en inspirerende periode ervaren. De constatering dat huidige eigen inzichten en vroegere standpunten niet altijd meer identiek zijn, doet daaraan niets af.

Diverse mensen hebben elk op hun eigen wijze aan de totstandkoming van dit proefschrift een bijdrage geleverd. Ik wil graag hen allen op deze plaats daarvoor bedanken.

In de eerste plaats wil ik mijn promotor en vroegere 'vakbaas' Prof. dr. L.F.W. de Klerk dankzeggen. Hij heeft mij indertijd mede gestimuleerd tot het doen van dit onderzoek. Bovendien heeft hij in de afgelopen jaren op gepaste wijze mij gemotiveerd tot afronding van de onderzoeksrapportage. Bij het schrijven heb ik graag gebruik gemaakt van zijn zeer snelle, kritische en opbouwende feedback.

Bij de opzet en uitvoering van het onderzoek heb ik veel steun ondervonden van drs. H. v. Huygevoort en drs. F. v.d. Schoot, toendertijd als student-assistenten aan het onderzoeksproject verbonden.

Als inhoudsdeskundige heeft drs. G. de Jong van het Talencentrum der KHT een belangrijke bijdrage geleverd aan het screenen van lessen en toetsen en het mede construeren van items. Ik ben hem daarvoor zeer erkentelijk.

Aan de proefpersonen die deelnamen aan de experimenten ben ik dank verschuldigd. Zonder hen zou dit onderzoek niet mogelijk zijn geweest.

Bij de verwerking van de onderzoeksgegevens heb ik veel steun gehad van het commentaar en de adviezen van drs. N. Veldhuijzen, medewerker van de afdeling OPD. Ook de heer P. Goldebeld, eveneens medewerker van de afdeling OPD, dank ik voor zijn enthousiaste en nauwgezette hulp achter de computer-

terminal. Dat ik beiden erkentelijk ben, zal hen bekend zijn, het is goed dit op deze plaats nog eens te benadrukken.

Prof. dr. E.J. Bijnen en prof. dr. E. Warries dank ik voor hun opbouwende commentaar op onderdelen van dit proefschrift.

Voor de typografische vormgeving van deze studie ben ik dank verschuldigd aan een groot aantal mensen. Het typen van het proefschrift werd verzorgd door mevr. A.E. Herrewijn-de Lavaletta en mevr. W.J. Kuysten-Speijers, secretaresses van de afdeling OPD, en mevr. D. Mulder-Vulkers, medewerkster van de afdeling Tekstverzorging. Mevr. M.M. van Dam-van der Velden ben ik erkentelijk voor de drukwerkvoorbereiding en de heer D. Verwoert voor het tekenen van de figuren.

Mevr. C.J. Theunissen-Burke dank ik voor haar vertaling van de samenvatting.

De Subfaculteit Psychologie van de KHT dank ik voor de beschikbaarstelling van gelden uit de onderzoeksbegroting, waardoor het onderzoek uitgevoerd kon worden. De directie van het Cito, bestaande uit drs. J.W. Solberg (directeur), drs. K.D. Thio en mr. W.F. van Dorp, ben ik dank verschuldigd voor de tijd die mij ter beschikking is gesteld voor het schrijven van dit proefschrift, voor het gebruik van de computer en voor de faciliteiten voor het drukken van de dissertatie. Tenslotte ben ik het Cito ook bijzonder erkentelijk voor de bijdrage in de druk- en uitgavekosten van dit werk.

Het voltooien van een dissertatie wordt vaak beschouwd als de afronding van een academische opleiding. Ik hecht er aan op dit moment en op deze plaats twee mensen in het bijzonder te danken voor hun rol die zij hierin hebben mogen spelen.

Allereerst wil ik mijn moeder danken voor de vele moeite die zij zich getroost heeft om mijn studie, eerst aan de Gem. Kweekschool te Rotterdam en later aan de Rijksuniversiteit Utrecht, mogelijk te maken. Ik ben haar daarvoor, meer dan deze zinnen kunnen uitdrukken, bijzonder erkentelijk.

Mijn vrouw, Winny, is een zeer belangrijke stimulans geweest voor de uiteindelijke totstandkoming van dit proefschrift. Reeds vanaf ons eerste, gezamenlijke begin aan de studie pedagogiek, is haar belangstelling voor mijn werk, haar kritische houding en haar emotionele steun voor mij zeer belangrijk geweest. Door het scheppen van de juiste sfeer werd het schrijven van deze dissertatie eerder een plezierig gebeuren dan een opgave. Ik hoop te zijner tijd voor haar een soortgelijke rol te mogen vervullen.

Peter Weeda

Maastricht, oktober 1982.

INHOUDSOPGAVE

Woord vooraf

1.	<i>Inleiding</i>	13
1.1.	Onderwijskunde en onderwijspsychologie	13
1.2.	Miniatuur-theorieën	14
1.3.	De opzet van deze studie	15
2.	<i>Individuele verschillen</i>	19
2.1.	Inleiding	19
2.2.	Individuele verschillen en afstemmingsmodellen	19
2.2.1.	Het onderwijsleermodel van Carroll	22
2.3.	De strategie van beheersingsleren	26
2.4.	De Keller-planbenadering	29
2.4.1.	Overeenkomsten en verschillen met de beheersingslerenstrategie	31
2.5.	Resultaten van empirisch onderzoek betreffende strategie van beheersingsleren	33
2.6.	Enige kanttekeningen bij de strategie van beheersingsleren	35
3.	<i>Een model voor het leren op school</i>	39
3.1.	Het model	39
3.2.	De leerling kenmerken	40
3.2.1.	Het cognitief begingedrag	40
3.2.2.	Affectieve leerlingkenmerken	44
3.3.	Kwaliteit van de instructie	46
3.3.1.	Feedback	47
3.3.2.	Correctieve procedures	49
3.4.	Leerresultaten	50
3.4.1.	Affectieve leerresultaten	51
3.4.2.	Leertempo	52
3.5.	Het model geïntegreerd	55
3.6.	Enige kritische kanttekeningen	56
4	<i>Vraagstellingen</i>	59
4.1.	Inleiding	59
4.2.	Vraagstellingen betreffende het model	60
4.3.	Vraagstellingen met betrekking tot de cognitieve leerresultaten	60
4.4.	Vraagstellingen met betrekking tot de affectieve modelcomponent	61
4.5.	Vraagstellingen met betrekking tot het leertempo (= de leertijd)	62
4.6.	Vraagstellingen betreffende de kwaliteit van de	

instructie	63
5. <i>Het gebruikte materiaal</i>	65
5.1. Inleiding	65
5.2. Inhoud en vormgeving van de cursus	65
5.2.1. Relevantie	65
5.2.2. Opzet van de cursus	66
5.2.3. Doelstellingen	68
5.2.4. Formatieve toetsen	68
5.2.5. Remediele instructie	71
6. <i>Psychometrische aspecten der formatieve toetsen</i>	73
6.1. Inleiding	73
6.2. Norm-georiënteerd en criterium-georiënteerd meten	73
6.3. Criterium-georiënteerde toets- en itemanalyse	75
6.4. Cesuurbepaling	76
6.4.1. Utiliteiten	78
6.5. Betrouwbaarheidsbepaling	80
6.5.1. De methode van Carver	81
6.5.2. De methode van Swaminathan, Hambleton & Algina	81
6.5.3. De methode van Huynh	81
6.5.3.1. Het bivariate beta-binomiale model	81
6.5.3.2. Berekening en evaluatie van κ (coëff. kappa)	83
6.5.4. De methode van Subkoviak	85
6.5.5. De methode van Marshall-Haertel	87
6.5.6. Vergelijking en evaluatie der methodes	87
6.6. Psychometrische analyse formatieve toetsen experiment I	88
6.6.1. Model-toetsing	88
6.6.2. Cesuurbepaling formatieve toetsen experiment I	92
6.6.3. Consistentiebepaling formatieve toetsen experiment I	96
6.7. Psychometrische analyse formatieve toetsen experiment II	99
6.7.1. Model-toetsing	99
6.7.2. Cesuurbepaling formatieve toetsen experiment II	101
6.7.3. Consistentiebepaling formatieve toetsen experiment II	103
7. <i>Experiment I: toetsing van het model</i>	105
7.1. Inleiding	105
7.1.1. Vraagstellingen betreffende het model	105
7.1.1.1. Hoofdeffecten	106
7.1.1.2. Interacties	107
7.1.2. De voorspellende waarde van het model na	

	correctie voor tijd	107
7.1.3.	Vraagstellingen betreffende feedback en correctieve procedures	108
7.2.	Proefpersonen	108
7.3.	Instrumenten voor de meting der leerling-kenmerken	110
7.3.1.	Instrumenten ten behoeve van het cognitieve begingedrag	111
7.3.2.	Instrumenten met betrekking tot de affectieve aanvangskenmerken	113
7.4.	Vormgeving van het onderwijsleerproces	115
7.4.1.	De leertaak	116
7.5.	Instrumenten voor de meting der leerresultaten	119
7.5.1.	Instrumenten ten behoeve van de cognitieve leerresultaten	120
7.5.2.	Meting van de leertijd/leertempo	120
7.5.2.1.	Enige kanttekeningen bij het meten van studietijden	121
7.5.3.	Meting van de affectieve leerresultaten	123
7.6.	Procedure	123
7.7.	Design	127
7.8.	Data-verwerking	127
7.9.	Resultaten	129
7.9.1.	Leerdoelgebonden toetsen	129
7.9.2.	Veranderingen in cognitieve leerprestaties over de leertaken	130
7.9.3.	Het cognitief begingedrag (CBG)	135
7.9.3.1.	Kanttekeningen bij de gehanteerde regressie-analyse techniek	136
7.9.3.2.	Voorspellende waarde van het cognitief begingedrag	137
7.9.4.	De affectieve leerlingkenmerken	145
7.9.5.	Kwaliteit van de instructie	149
7.9.5.1.	Descriptieve gegevens	149
7.9.5.2.	Hoofdeffecten	149
7.9.5.3.	Variaties in feedback en correctie	151
7.9.6.	Het totale model	156
7.10.	Discussie	158
8.	<i>Experiment II: een type 3 onderzoek</i>	
8.1.	Inleiding	163
8.2.	Hypotheses	165
8.3.	Proefpersonen	166
8.4.	Instrumenten voor de meting van het cognitief begingedrag	168
8.5.	Vormgeving van het onderwijsleerproces	169

8.6.	Instrumenten voor de meting der cognitieve leerresultaten	170
8.7.	Meting van de leertijd	171
8.8.	Procedure	171
8.9.	Design en data-analyse	173
8.10.	Resultaten	174
8.10.1.	Descriptieve gegevens	174
8.10.2.	Toetsing der hypothesen	176
8.10.3.	Toetsing der hypothesen na correctie voor leertijd	180
8.11.	Discussie	181
9.	<i>Algemene discussie</i>	185
9.1.	De belangrijkste onderzoeksresultaten	185
9.2.	Theoretische implicaties	188
9.2.1.	Inleiding	188
9.2.2.	Beheersingsleren en ATI-onderzoek	189
9.2.3.	Beheersingsleren en het 'direct-teaching model'	192
9.3.	Praktische implicaties	193
9.4.	Suggesties voor verder onderzoek	194
	Samenvatting	197
	Summary	203
	Literatuur	209

Hoofdstuk 1

INLEIDING

1.1. Onderwijskunde en onderwijspsychologie

De onderwijskunde is een relatief jonge wetenschap. De aanduiding van een discipline met deze naam is de facto pas in het midden der zestiger jaren ontstaan. Kenmerkend voor jonge wetenschappen is vaak de zorg om de eigen identiteit. In de beginjaren zeventig is een tweetal conferenties volledig gewijd aan de formulering wat onder onderwijskunde begrepen moet worden. Als omschrijving van het vakgebied vond men elkaar op de volgende formulering: 'Onderwijskunde is gericht op wetenschappelijk verantwoorde bestudering en analyse van het onderwijs in al zijn facetten om, mede op basis daarvan bij te dragen tot de oplossing van onderwijsvraagstukken (Rapport, 1973). Ook in de bekende inleiding in de onderwijskunde 'Beknopte didaxologie' (De Corte et al., 1981) wordt bij deze omschrijving aangesloten, alhoewel niet gesproken wordt over *oplossing*, maar over *optimalisering* van structuren en processen (o.c. p. 3-4).

De onderwijskunde wortelt qua ontstaan in de pedagogiek en - in mindere mate - in de psychologie. De benaderingswijze van de pedagogiek was daarbij sterk normatief van aard, terwijl de psychologie zich voornamelijk bezighield met het leren en leermoeilijkheden. Allengs vond echter terreinuitbreiding plaats en werd de gehanteerde methode van onderzoek meer verruimd en meer empirisch van karakter (zie bijvoorbeeld Van Trotsenburg, 1972).

De onderwijskunde begon zich tevens langzamerhand te onttrekken aan de verschillende basiswetenschappen en ontwikkelde zich tot een - binnen zekere grenzen - zelfstandige discipline.

Naast een ontwikkeling gericht op een geïntegreerde zelfstandige discipline onderwijskunde, waarin een interdisciplinaire werkwijze wordt aangehangen, zien we ook een ontwikkeling naar een verzelfstandiging van diverse oorspronkelijke basis-wetenschappen, waarin een monodisciplinaire aanpak wordt voorgestaan. Zo zijn ontstaan de onderwijspsychologie en de onderwijssociologie. Merkwaardig is in dit verband dat de monodiscipline schoolpedagogiek in ieder geval als begrip is verdwenen, dan wel is vervangen door de term (en studierichting) onderwijskunde.

We kunnen twee soorten onderwijskunde onderscheiden. Het ene type onderwijskunde is de wetenschap die uitgaat van een meer interdisciplinaire aanpak en een eigensoortige bijdrage aan de

theorievorming. Het tweede type onderwijskunde is meer een verzamelnaam voor verschillende monodisciplines, die zich elk met een eigen wetenschappelijke methode, richten op (deelgebieden van) het onderwijsgebeuren.

De Klerk (1974) bedoelt de onderwijskunde van het tweede type als hij spreekt over 'De geordende kennis die door middel van de diverse wetenschappelijke activiteiten wordt bijeengebracht, wordt wel aangeduid met de term onderwijskunde' (o.c. p.3).

De onderwijspsychologie kan nu gerekend worden tot de onderwijskunde van het tweede type. Zij onderscheidt zich van de vroegere schoolpsychologie door de concentratie op onderwijsleerprocessen en van de leerpsychologie door het optimaliseringsgezichtspunt (cf. Van Parreren 1981, p.120). Naast de term onderwijspsychologie wordt ook wel gesproken van onderwijsproceskunde (Van Parreren, 1981) en onderwijsleerpsychologie (Span, 1982). De term onderwijspsychologie verdient de voorkeur daar de monodisciplinaire gerichtheid daarbij het duidelijkst tot uitdrukking komt.

De onderhavige studie kan naar de aard van zijn vraagstelling en wijze van theorievorming gerekend worden tot de onderwijskunde van het eerste type.

1.2. *Miniatuur-theorieën*

De Klerk (1982) stelt dat algemene psychologische theorieën meestal weinig zinvol zijn voor het afleiden van specifieke hypothesen over de aard van de samenhang tussen de resultaten van het onderwijs, het type leerlingen en de gehanteerde onderwijsmaatregelen in concrete onderwijsleersituaties. Hij pleit in dit verband voor *miniatuurtheorieën* die slechts betrekking hebben op bepaalde aspecten van onderwijsleersituaties. Dergelijke theorieën dienen tenminste prescriptief (cf. Bruner, 1966) te zijn. Dat wil zeggen uit de theorie moet blijken op welke wijze gegeven doelstellingen door de leerlingen zo goed mogelijk bereikt kunnen worden. Overigens dient hierbij wel bedacht te worden, dat leerlingen langs verschillende wegen de doelstellingen kunnen bereiken. Binnen de theorie dient dan ook rekening gehouden te worden met dergelijke individuele verschillen.

De in dit proefschrift nader onderzochte theorie van Bloom betreffende het leren op school is geen miniatuur-theorie. Wel biedt dit model diverse aanknopingspunten voor de ontwikkeling van miniatuur-theorieën. Door ons is met name aandacht besteed aan de relatie tussen bepaalde onderwijskenmerken en leerlingvariabelen.

Er kunnen grofweg drie soorten onderwijskenmerken onderscheiden worden:

1. pre-instructie-kenmerken, hierbij gaat het met name om de invloed van pre-instructievariabelen (bijvoorbeeld 'advance organizers', doelstellingen) op de onderwijsresultaten;
2. instructie-kenmerken, hierbij gaat het om de invloed van instructievariabelen (bijvoorbeeld structuur van de leerstof, aanwezigheid van plaatjes en schema's) op de onderwijsresultaten.
3. post-instructie-kenmerken, hierbij gaat het om de invloed van post-instructievariabelen (bijvoorbeeld feedback, correctieve procedures) op de onderwijsresultaten.

De onderwijskenmerken waaraan in deze studie aandacht is besteed behoren tot de categorie post-instructie-kenmerken. Geen aandacht is besteed aan de wijze waarop de diverse kenmerken hun invloed op de onderwijsresultaten doen gelden. Intermediërende variabelen zoals informatie-verwerkingsprocessen en cognitieve structuren kunnen daarbij een rol spelen. In deze studie wordt daarop niet ingegaan. Hiermee is tevens een beperking van het onderhavige onderzoek aangegeven.

1.3. De opzet van deze studie

Het hoofddoel van het in dit proefschrift beschreven onderzoek is het toetsen van het model van Bloom betreffende het leren op school. Het model pretendeert universeel te zijn, dat wil zeggen geldig voor diverse vormen van (schools)leren en voor diverse typen leerlingen. Gegeven deze pretenties kan moeilijk gesproken worden over een miniatuurtheorie. Dat kan wel waar het onderzoek naar de aard van de correctieve procedures betreft.

In hoofdstuk 2 wordt aandacht besteed aan de individuele verschillen die er tussen leerlingen bestaan. Aangegeven wordt welke de onderwijskundige mogelijkheden zijn voor aanpassing aan dergelijke verschillen. Vervolgens wordt aandacht besteed aan het Carroll-model waarin het tijdsaspect een belangrijke plaats inneemt. Tenslotte wordt uitvoerig ingegaan op de strategie van beheersingsleren, die in het eerder genoemde model van Bloom een centrale rol vervult.

In hoofdstuk 3 wordt het model van Bloom betreffende het leren op school besproken. Aan de inputkant van het model staan de cognitieve en affectieve leerlingkenmerken. Het procesgedeelte van het model wordt gewijd aan de kwaliteit van de instructie. De output-kant tenslotte wordt gevormd door de cognitieve leerresultaten, de affectieve resultaten en het leertempo. Het vierde hoofdstuk bevat de diverse vraagstellingen die de uit-

gangspunten vormen van het in de latere hoofdstukken beschreven door ons uitgevoerde onderzoek. Bloom geeft, op basis van een aanzienlijke hoeveelheid empirisch onderzoek, aan hoeveel procent van de variantie in leerprestaties door de diverse componenten van zijn model kan worden verklaard. Deze claims vormen in feite de vraagstellingen van ons onderzoek. Hoofdstuk 5 geeft een beschrijving van het door ons gebruikte onderwijsmateriaal. De cursus heeft betrekking op het leren van diverse grammaticale regels in het Frans. Het is geen cursus bedoeld voor hen die voor het eerst Frans leren, maar voor hen die reeds een redelijke kennis van de Franse taal hebben.

Het zesde hoofdstuk is geheel gewijd aan de psychometrische analyse van de diverse door ons in het onderzoek gebruikte formatieve toetsen. Aandacht wordt geschonken aan de beperkingen die de klassieke testtheorie voor dergelijke toetsen heeft en aan oplossingen die voor deze beperkingen ontwikkeld zijn.

Hoofdstuk 7 vormt het hoofdgedeelte van dit proefschrift. Het geeft een beschrijving van het door ons uitgevoerde onderzoek betreffende het model van Bloom voor het leren op school, waarbij de strategie van beheersingsleren centraal staat. Enerzijds wordt nagegaan of het gedrag van de leerlingen overeenkomstig de modelvoorspellingen is en anderzijds wordt nagegaan welke de variantie-verklarende waarde van de verschillende modelcomponenten is. Een betrekkelijk nieuw aspect in ons onderzoek met betrekking tot het beheersingsleren wordt gevormd door de geregelde meting van de leertijd der leerlingen. Immers het gaat niet aan de superioriteit van de ene onderwijsmethode boven een andere aan te tonen, zonder rekening te houden met het feit dat de eerstgenoemde aanmerkelijk meer leertijd in beslag neemt dan de laatstgenoemde. Het achtste hoofdstuk bevat het verslag van een onderzoek waarin twee correctieve procedures met elkaar vergeleken worden. Tevens wordt daarbij nagegaan of leerlingen met verschillend cognitief begingedrag verschillen in hun leerprestaties bij de onderscheiden correctieve procedures. Ook nu wordt wederom rekening gehouden met de leertijd en worden onder meer efficiëntiescores berekend.

Het negende en laatste hoofdstuk tenslotte bevat een uitvoerige discussie waarin enerzijds aangegeven wordt welke de wetenschappelijke implicaties zijn van het uitgevoerde onderzoek, terwijl anderzijds ook ingegaan wordt op de praktische implicaties van de experimenten. Voorts worden de onderzoeksresultaten ook gerelateerd aan literatuur betreffende eerder verricht empirisch onderzoek.

Tenslotte wordt nog aandacht besteed aan de diverse

beperingen waarvan onvermijdelijk elk onderzoek, dus ook het onderhavige, te lijden heeft.

Hoofdstuk 2

INDIVIDUELE VERSCHILLEN

2.1. Inleiding

Het is een gegeven dat niet alle mensen gelijk zijn. Evenzo zijn niet alle leerlingen gelijk. Op zichzelf is dit geen schokkende constatering, ware het niet dat individuele verschillen tussen leerlingen implicaties hebben voor het aan die leerlingen te geven onderricht. Velen hebben zich dan ook al beziggehouden met de vraag op welke wijze men in het onderwijs rekening kan houden met individuele verschillen tussen de leerlingen (zie voor een beknopt historisch overzicht bijvoorbeeld Nijhof, 1978, p.12-52). Glaser (1972) omschrijft individualisering in het onderwijs als het aanpassen van het onderwijs aan individuele verschillen tussen leerlingen. Weisgerber (1972) kiest in het kader van het project PLAN (Program for Learning in Accordance with Needs) voor de term 'individualized learning' in plaats van 'individualized instruction'. Weisgerber stelt de leerling centraal, terwijl Glaser ervan uitgaat dat er een samenhang bestaat tussen onderwijsvormen en leerlingkenmerken zoals aanleg en studietempo. De benadering van Glaser is wat realistischer dan die van Weisgerber. Immers, hij houdt ook rekening met de eventuele beperkingen van het onderwijsstelsel, waardoor implementatie van onderwijsveranderingen gemakkelijker kan geschieden.

In dit hoofdstuk zal kort ingegaan worden op een van Cronbach (1967) afkomstige indeling van afstemmingsmogelijkheden van het onderwijs op individuele verschillen tussen leerlingen. Vervolgens zal meer in het bijzonder aandacht besteed worden aan de strategie van beheersingsleren en het Keller-plan. Beide vormen van onderwijsarrangementen waarin rekening wordt gehouden met individuele verschillen tussen leerlingen én met de realiseringmogelijkheden die de gangbare praktijk biedt. Beide voldoen dus aan de definitie van Glaser.

2.2. Individuele verschillen en afstemmingsmodellen

Het verloop en de resultaten van onderwijsleerprocessen (kunnen) worden beïnvloed door persoonlijke en sociale gegevens van de kant van leerlingen en docenten, alsmede door overige niet-persoonsgebonden situationele gegevens. Het verdient aanbeveling om bij de planning van onderwijsleerprocessen rekening te houden met het conglomeraat van deze

factoren. Met andere woorden: bij de planning zal mede in ogenschouw genomen moeten worden dat er zogenaamde inter-individuele verschillen bestaan tussen de leerlingen, die mogelijkwijze gevolg(en) kunnen hebben voor de opzet, uitvoering en uitkomsten van onderwijsleerprocessen.

Het onderkennen van interindividuele verschillen tussen de leerlingen (wij beperken ons in dit verband tot de leerlingen) leidt tot de wens om aan deze verschillen aangepaste, dus optimale onderwijsarrangementen ('instructional designs') te ontwerpen. Onder een onderwijsarrangement wordt verstaan een geheel van onderwijsprocedures en methoden, dat erop gericht is de actuele (begin)toestand van de leerling te wijzigen in de gewenste (eind)toestand (De Klerk, 1979). De aanpassing van onderwijsarrangementen kan op velerlei wijzen gebeuren. Cronbach (1967, p. 24 e.v.) verdeelt de aanpassingsvormen naar twee aspecten: het leerdoel en de instructievorm. Op deze wijze ontstaan een vijftal afstemmingsmodellen (Tabel 2.1.).

Tabel 2.1. Onderwijskundige afstemmingsmodellen voor interindividuele verschillen

leerdoelen	instructie- wijze	adaptaties a/d inter- individuele verschillen
vast	vast	1a. Varieer de schoolperiode d.m.v. sequentiële selectie. 1b. Ga door met instructie tot het criterium bereikt is, m.a.w. verleng de instructietijd.
optioneel	vast binnen een optie	2. Bepaal voor elke leerling de voor hem relevante doelen en maak daarvoor een curriculum.
vast binnen een programma (onderdeel)	alternatieven aanbieden	3a. Lever remediële aanvullingen aan de vaste hoofdlijn van de instructie. 3b. Instrueer verschillende leerlingen met verschillende methoden.

Het bekendste afstemmingsmodel is naar alle waarschijnlijkheid het zogenaamde sequentiële selectiemodel. Het vigerende Nederlandse onderwijssysteem is daarvan (helaas) nog steeds een voorbeeld. Steeds meer raakt men echter overtuigd van de beperkingen van dit model en zoekt men naar mogelijkheden tot individualisering. Vaak neemt dit de vorm aan van aanpassing van het doel aan de individuele leerlingen. Dit heeft onder meer geleid tot het onvoldoende tot ontwikkeling komen van effectieve instructiewijzen, of tot onvoldoende aandacht hiervoor. Het in Tabel 1, onder 2a beschreven afstemmingsmodel is een voorbeeld van een dergelijke benaderingswijze. Het kan leiden tot aanzienlijke verstarring binnen het onderwijssysteem, terwijl een zware wissel getrokken wordt op het selectiesysteem. Betere mogelijkheden worden geboden door de onder 3a en 3b genoemde afstemmingsmodellen, waarin de verschillende instructiewijzen centraal staan.

Het onder 3a genoemde afstemmingsmodel is op grond van overwegingen van meer praktische aard in eerste instantie het aantrekkelijkst. Immers de gekozen instructieprocedure is voor alle leerlingen unifom, behalve voor die leerlingen waarbij geconstateerd wordt dat na doorlopen van het onderwijsleerproces het voorafgestelde leercriterium (nog) niet is bereikt. Zij krijgen dan aanvullende, zogenaamde remediële instructie. Met andere woorden zij volgen de een of andere *correctieve procedure*. De aard en vorm der correctieve procedures zijn weer voor alle leerlingen die daarvoor in aanmerking komen gelijk. Onderzoeksvragen in dit verband hebben betrekking op de vraag welke correctieve procedures nu het meeste rendement opleveren. Deze vraag komt in het kader van het onderwerp van deze studie aan de orde in Hoofdstuk 4. Het onder 3b genoemde afstemmingsmodel impliceert onder meer dat er een eenduidige procedure is op grond waarvan leerlingen aan alternatieve instructieprocedures toebedeeld kunnen worden.

Sterke nadruk komt hierbij te liggen op de keuze van de variabele(n) waarover informatie verzameld wordt, bedoeld voor de toewijzing van leerlingen aan alternatieve instructieprocedures. Dit type onderzoek wordt wel aangeduid met de benaming 'aptitude-treatment interaction research'.

Een door Cronbach niet genoemd afstemmingsmodel is het model dat gevormd wordt door de combinatie van de modellen 3a en 3b. Alle leerlingen ondergaan in eerste instantie dezelfde instructieprocedure. Vervolgens komen leerlingen die het vereiste leercriterium niet bereikt hebben in aanmerking voor aanvullende remediële instructie, echter niet voor alle leerlingen dezelfde remediële instructie. Men kan vanaf dit moment gaan werken met alternatieve correctieve procedures.

2.2.1. Het onderwijsleermodel van Carroll

In 1963 publiceerde John B. Carroll een artikel waarin hij een conceptueel model beschreef betreffende factoren die van invloed zijn op succes in het schoolse, groepsgewijze leren. Het is een typisch onderwijskundig model, nauwelijks verwant aan psychologische theorieën die zich met het leren van individuen onder gecontroleerde omstandigheden bezighouden (Warries, 1979, p.11). In zijn eenvoudigste vorm stelt het model dat de beheersingsgraad van een leertaak door een leerling afhankelijk is van de verhouding tussen de tijd die de leerling eraan heeft besteed en de tijd die hij nodig zou hebben om de taak af te maken.

Carroll ging ervan uit dat in de schoolsituatie de door de leerling bestede en benodigde tijd niet alleen wordt beïnvloed door karakteristieken van de leerling, maar ook door instructiekenmerken.

De bestede tijd ('time spent') wordt bepaald door de *leermotivatie* ('perseverance') en de *roostertijd* ('opportunity to learn'), terwijl de benodigde tijd bepaald wordt door het *specifieke leertempo* van de leerling ('aptitude'), de *kwaliteit van het onderwijs* ('quality of instruction') en het *begripsvermogen* van de leerling ('ability to understand instruction').

De vijf modelfactoren worden door Carroll tamelijk globaal gedefinieerd. De leermotivatie wordt omschreven als de hoeveelheid tijd die de leerling actief met leren wil doorbrengen. De roostertijd wordt gedefinieerd als de door de school, of de leerkracht ter beschikking gestelde tijd (bijvoorbeeld de leerlingen krijgen maximaal 20 minuten de tijd om een bepaald wiskundig probleem op te lossen). Het specifiek leertempo van de leerling ('aptitude') wordt omschreven als de tijd die de leerling nodig heeft om een bepaalde leertaak onder ideale onderwijscondities te volbrengen. Het is derhalve een, niet perse noodzakelijk, constante eigenschap van de leerling, maar veeleer afhankelijk van de leertaak. De kwaliteit van het onderwijs, alsmede het begripsvermogen van de leerling tenslotte bepalen in onderlinge wisselwerking de extra tijd die de leerling nodig heeft om de leertaak te beheersen onder niet-optimale condities. Als de kwaliteit van het instructieproces erg hoog is, dan zal de leerling snel de leertaak beheersen en slechts geringe extra tijd nodig hebben dan die overeenkomstig zijn specifiek leertempo noodzakelijk is. Indien de leerling echter te maken heeft met een kwalitatief matig instructieproces, dan zal hij relatief veel moeite hebben met het beheersen van de leertaak en dus veel extra tijd nodig hebben. Het onderwijsleermodel van Carroll

kan als volgt voorgesteld worden (cf. Block en Burns, 1977, p.6):

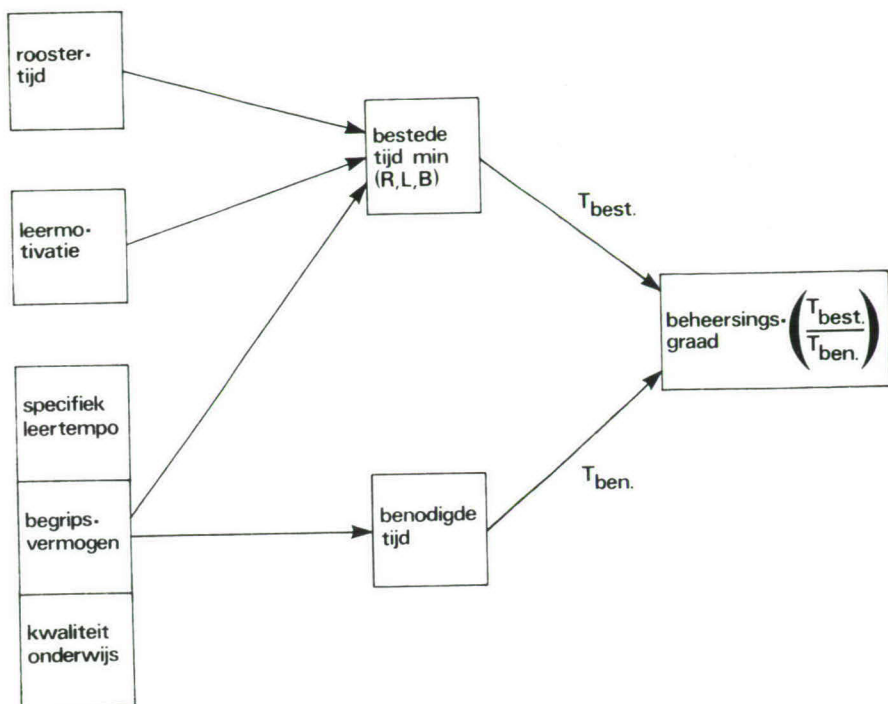
$$\begin{aligned} \text{beheersingsgraad} &= f \left[\frac{\text{bestede tijd}}{\text{benodigde tijd}} \right] \\ &= f \left[\frac{\text{leermotivatatie of roostertijd}}{\text{benodigde tijd (leertaak) + benodigde tijd (instructie)}} \right] \\ &= f \left[\frac{\text{leermotivatatie of roostertijd}}{\text{specifiekleertempo + kwaliteit onderwijs X begripsvermogen}} \right] \end{aligned}$$

Bovenstaande 'formule' geeft aan dat de beheersingsgraad van een bepaalde leertaak afhankelijk is van de verhouding tussen de leermotivatatie of roostertijd enerzijds en het specifiek leertempo, de kwaliteit van het onderwijs en het begripsvermogen anderzijds. Het lijkt ons terminologisch onjuist om het linkerlid van bovenstaande formule, te weten beheersingsgraad, gelijk te stellen aan criteriumscore zoals Warries (1979a, p.15) doet. Immers de beheersingsgraad van een leertaak kan variëren, en is ook afhankelijk van de leerling. De criteriumscore daarentegen is een door de docent vastgestelde score die voor een bepaalde leertaak constant is (bijv. de criteriumscore is 8 items goed van een leerdoelgerichte toets bestaande uit 10 items).

Duidelijker dan in bovenstaande formules worden de relaties tussen de diverse modelfactoren weergegeven in fig. 2.1. (Cooley en Lohnes, 1976, p.188). Aangezien Carroll alle variabelen gedefinieerd heeft in termen van tijd kan de beheersingsgraad nu berekend worden.

Cooley en Lohnes (1976, p.188) merken op dat hoewel Carroll zelf in eerste instantie meende dat de variabele 'kwaliteit van het onderwijs' nogal vaag was ('the most elusive quantity in this model', Carroll, 1963, p. 829), een dergelijk oordeel ook geveld zou kunnen worden aangaande de variabele 'begripsvermogen'. De enige aanwijzing voor een nadere begripsomschrijving vormt de idee dat de variabele verschilt van het specifiek leertempo en dat het begripsvermogen waarschijnlijk een combinatie is van algemene intelligentie en verbale bekwaamheid.

Voor wat betreft de variabele 'kwaliteit van het onderwijs' geeft Carroll (o.c.) een drietal condities aan die kwaliteitsverhogend kunnen werken, te weten:



Figuur 2.1. Carroll's onderwijsleermodel

1. de leerling moet verbaal duidelijk worden gemaakt dat hij kan leren wat hij moet leren en op welke wijze hij het moet leren;
2. de leerling moet op zinvolle wijze in contact gebracht worden met het leermateriaal;
3. de verschillende onderdelen van de leertaak moeten in een zodanige volgorde en in een zodanige mate van gedetailleerdheid aangeboden worden, dat zoveel mogelijk elke leerstap adequaat voorbereid is door een voorafgaande leerstap (p. 726).

De door Carroll aangeduide condities zijn nogal vaag en algemeen van aard. Zij bieden geen concrete aangrijpingspunten voor de constructie van onderwijsarrangementen. In algemene zin verwijzen zij echter wel naar allerlei theorieën en modellen waarin meer uitgewerkte aanwijzingen voor de inrichting van het onderwijsleerproces zijn te vinden. Met name de opvattingen van Glaser (1972) aangaande de relatie tussen het actuele begingedrag van de leerling en het vereiste begingedrag en de opvattingen van Gagné (1970, 1974) met betrekking tot de toepassing van allerlei leertheoretische

principes bij de vormgeving van de onderwijsarrangementen kunnen aanknopingspunten bieden voor optimalisering van de variabele 'kwaliteit van het onderwijs'. Los van de impliciete vaagheid van de theoretische descriptie zoals die door het model van Carroll gepresenteerd wordt, hetgeen overigens ook in andere instructiemodellen en -theorieën - bijvoorbeeld Ausubel 1968 -voorkomt, is het uitgangspunt dat het leren als proces univariaat is wellicht problematisch.

Immers Carroll operationaliseert in zijn model 'aptitude' als de tijd die een leerling nodig heeft om een bepaalde leertaak te beheersen. Wanneer men er nu vanuit gaat dat een leerling 'het criterium bereikt' op verschillende tijdstippen voor verschillende variabelen (tezamen een leertaak vormend), dan gaat men uit van een multivariate opvatting met betrekking tot het leerproces. Een dergelijke opvatting heeft onder meer consequenties voor de mogelijk geachte leercurves. Indien er immers verschillende leercurves ontstaan, dan kunnen deze niet door middel van één dimensie (leertijd) verklaard worden.

Men kan zich afvragen in hoeverre het model van Carroll correct is door individuele verschillen tussen leerlingen terug te brengen tot verschillen in leertijd. Cronbach & Snow (1977, p. 120) melden, dat uit een persoonlijke mededeling van Carroll blijkt, dat laatstgenoemde meent dat het model met name van toepassing is op het leren van vreemde talen, omdat uit de ervaring blijkt dat de verschillende leervorderingen elkaar stapsgewijs opvolgen als goede, gevarieerde onderwijsmethoden gebruikt worden.

Cronbach & Snow (1977, p.120) stellen dat het construct leersnelheid slechts van beperkte waarde is als het alleen geldig is onder dergelijke specifieke condities.

Hiertegen kan men echter aanvoeren, dat het ontstaan van deel- of miniatuur-theorieën binnen de sociale wetenschappen na verloop van tijd aanleiding kan vormen voor integratie van een en ander in een meer omvattende algemene theorie. Vanzelfsprekend zal men, bij gebruikmaking van een algemene groeimaat, zoals leertijd/leersnelheid, zich moeten realiseren dat men - later - ook de leercurves voor de diverse (sub)leerresultaten zal moeten analyseren en verklaren. Het is evident dat er geen 'overall criterion' is waarop het onderwijsleerproces zich kan richten. Voor onderzoeksdoeleinden kan men echter voorlopig wel met een dergelijk beperkt gezichtspunt volstaan, waarbij secundaire analyses van onderzoeksresultaten misschien ook antwoorden kunnen verschaffen op de meer gedetailleerde vragen. Voorlopig kan het model van Carroll nog beschouwd worden als een heuristisch model dat aanleiding kan

geven tot vele onderzoeksvragen en tot onderwijskundige theorievorming.

2.3. De strategie van beheersingsleren

De strategie van beheersingsleren ('mastery learning'), opgesteld en uitgewerkt door Benjamin S. Bloom (1968), is gebaseerd op het in de vorige paragraaf beschreven model van Carroll. De strategie is in de concrete onderwijspraktijk op diverse manieren uitgewerkt. Binnen deze verscheidenheid aan uitwerkingen zijn overeenkomstige factoren aan te wijzen. Warries (1979b, p.4) geeft een *minimale definitie* in vijf punten van de beheersingsleren-strategie, t.w.:

1. De einddoelen van 'de cursus' liggen vast.
2. De leerinhoud (de leerstof) is opgedeeld in leereenheden of leertaken; een leerling moet voorgaande taken beheersen.
3. Na elke leertaak volgt toetsing en snelle informatie over de uitslag.
4. Bij onvoldoende beheersing van een taak is correctief studiemateriaal aanwezig.
5. Er is telkens enige tempodifferentiatie, maar geen niveau-differentiatie.

Het formuleren van de einddoelen van de cursus en van die van de verschillende leertaken is geen aanbeveling die uitsluitend geldt voor de beheersingsleren-strategie. In het algemeen wordt aanbevolen de doelstellingen bij het ontwerpen van onderwijsarrangementen zo concreet mogelijk te expliciteren (zie o.a. Gagné & Briggs, 1974 en Möller, 1973). Bij systematische toepassing van de beheersingsleren-strategie is het in ieder geval een vereiste dat de gevraagde eindgedragingen goed gedefinieerd zijn. Pas daardoor wordt het mogelijk om ook de diverse (formatieve) toetsen te construeren.

Het beheersen van de leertaken is een essentieel kenmerk van de strategie. Vóórdat de leerling verder gaat in het onderwijsleerproces, moet het bewijs geleverd worden dat hij de voorafgaande leertaak beheerst. Duidelijk is in dit verband dat de leertaken niet te groot mogen zijn, daar anders de vertraging of de achterstand in het leerproces te groot kan worden. Over de toegestane omvang van een leertaak lopen de meningen uiteen. Warries (1979b, p.5) schat de omvang van een leertaak in het voortgezet onderwijs op plm. 10 à 20 uur, inclusief huiswerk en lessen. Zelf achten wij een dergelijke omvang te groot. Ook in het gerapporteerde empirisch onderzoek met betrekking tot de strategie voor beheersingsleren (zie bijvoorbeeld Block, 1974) is de gemiddelde omvang van een

leertaak beduidend geringer. In het in de hoofdstukken 7 en 8 van deze studie beschreven onderzoek is de omvang van een leertaak gelijkgesteld aan ongeveer één lesuur. Op deze wijze kan ons inziens het leerproces optimaal gestuurd worden zonder dat de daarvoor benodigde toetstijd te groot wordt.

Na elke leertaak wordt een zogenaamde formatieve toets (zie Hoofdstuk 6) afgenomen en krijgt de leerling informatie aangaande de mate waarin de leerstof beheerst wordt. De toets is formatief, hetgeen impliceert dat deze voor geen ander dan een informerend doel gebruikt mag worden. Het is dus niet de bedoeling de leerling te beoordelen (ten behoeve van een rapportcijfer of iets dergelijks) op grond van de op de formatieve toets behaalde resultaten.

Bloom, Hastings & Madaus (1971) gaan zeer uitvoerig en gedetailleerd in op de constructie van dergelijke formatieve toetsen. De basis van alle toetsconstructie wordt gevormd door de leerstofanalyse: het opstellen van de toetsmatrijs. Langs de ene as worden de leerstofonderdelen afgezet en langs de andere as de gedragscategorieën. Deze laatste geven aan welk waarneembaar gedrag de leerling ten aanzien van de diverse leerstofcomponenten moet demonstreren.

Een essentieel onderdeel in de strategie van beheersingsleren wordt gevormd door feedback en correctie. Het heeft slechts zin door middel van de formatieve toetsen hiaten in het leerproces aan te tonen, als daarna de mogelijkheid geboden wordt deze lacunes weg te werken. Hiervoor zijn de correctieve procedures bedoeld. Deze kunnen variëren van simpelweg herhalen van de totale leerstof, of van leerstofonderdelen, tot het bestuderen van alternatieve leerstof. In ons onderzoek (zie o.a. Hoofdstuk 5) hebben wij gemeend om met betrekking tot de correctiva zo dicht mogelijk aan te moeten sluiten bij de oorspronkelijke leerstof.

Het laatste punt uit de definitie van de beheersingslerenstrategie tenslotte heeft betrekking op een vorm van interne differentiatie. Met interne differentiatie wordt een bepaalde organisatie van het onderwijs aan een groep of klas bedoeld, waardoor de leerkracht in staat wordt gesteld dit onderwijs op de individuele leerling af te stemmen (zie bijvoorbeeld Nijhof, 1978). De mastery-learning-strategie maakt wel tempodifferentiatie, maar geen niveaudifferentiatie mogelijk. Bij tempodifferentiatie is het de bedoeling dat de leerlingen dezelfde onderwijsdoelstellingen bereiken zij het in een verschillend tempo. Doelmatige tijdsdifferentiatie moet er, volgens Warries (1979b, p.6) toe leiden, dat alle leerlingen die daaraan behoefte hebben, extra kansen worden geboden om hetzelfde eindniveau te behalen, dat eerst door weinigen werd behaald. Het aspect van de tempodifferentiatie is nogal eens

een bron van misverstand geweest zowel ten aanzien van het Carroll-model, als ten aanzien van de strategie van beheersingsleren. In al te populaire vorm werd vaak gesteld dat men alle leerlingen, ongeacht hun leercapaciteiten, alles kon leren. Zo werkt het natuurlijk niet. Wel is het zo dat de grote verschillen in leertijd die er in een gemiddelde schoolklas tussen de snelle en minder snelle leerlingen bestaan aanmerkelijk gereduceerd kunnen worden. Bloom spreekt over een verhouding in benodigde leertijd van 1 : 5 vóórdat en van 1 : 1,5 nadat de beheersingsleren-strategie is toegepast, (Bloom, 1974). En bovendien dat meer leerlingen dan voorheen het criterium kunnen behalen.

De differentiatie bij de strategie voor beheersingsleren heeft betrekking op de voortgang van de leerlingen na afname en scoring van de formatieve toets. Voor verschillende leerlingen worden dan verschillende beslissingen genomen. Er worden beslissingen genomen aangaande de mate van beheersing van de leerstof door de leerling; aangaande de vraag of de gehele leerstof, of onderdelen opnieuw bestudeerd moeten worden en tenslotte op welke wijze de leerstof opnieuw bestudeerd moet worden. Al deze voortgangsbeslissingen worden genomen door de leerkracht. Deze sterke mate van sturing door de leerkracht is kenmerkend voor beheersingsleren.

Het centrale begrip binnen het Carroll-model en Bloom's beheersingslerenstrategie is 'tijd'. Als in een onderwijsarrangement zoveel tijd aan individuele leerlingen ter beschikking wordt gesteld als zij nodig hebben, dan zal een zeer groot gedeelte der leerlingen (75 à 80%) een hoog niveau op de criteriumvariabele behalen. In feite heeft Bloom verschillende ideeën uit Carroll's model van het schoolse leren vertaald in een soort filosofie en technologie van leren en onderwijzen. Bloom beschouwt aanleg ('aptitude') niet als een voorspeller van uiteindelijk leersucces onder constant-houding van de leertijd. Veeleer beschouwt hij aanleg als een voorspeller van de tijd nodig om een bepaald leercriterium te halen in een ideaal onderwijsarrangement. Met andere woorden, individuele verschillen in aanleg worden conceptueel getransformeerd tot individuele verschillen in benodigde leertijd. In plaats van toe te staan dat individuele leerprestaties variëren, waarbij tijd en instructie constant worden gehouden, worden in de strategie van beheersingsleren tijd en instructie toegestaan over individuen te variëren, waarbij de leerprestaties constant gehouden worden. Dit principe wordt in figuur 2.2., ontleend aan Buss & Poley (1976), in beeld gebracht.

LEERPRESTATIES	TIJD	
	variabel	constant
	constant variabel	
	maximaliseer potentieel	traditioneel onderwijs
	beheersings- leren (aanvankelijk)	beheersings- leren (later)

Figuur 2.2. Vier onderwijsarrangementen gegenereerd door twee randvariabelen.

In figuur 2.2. worden vier mogelijke onderwijsarrangementen geschetst, ontstaan door het variëren of constant houden van respectievelijk de leertijd en de leerprestaties. In het traditionele onderwijs wordt de leertijd constant gehouden en variëren de leerprestaties, terwijl binnen het aanvankelijk beheersingsleren de te behalen leerprestaties constant blijven en de benodigde leertijd varieert. In een latere fase van het beheersingsleren stelt Bloom (1973, p.57) dat de individuele verschillen in tijd benodigd om het prestatie criterium te behalen langzamerhand verdwijnen. Zij naderen een 'vanishing point'. De leertijd wordt derhalve weer constant. Dit zou volgens Bloom binnen de bestaande onderwijskaders het ideale onderwijsarrangement zijn. De laatste cel in figuur 2.2. tenslotte geeft aan, dat zowel de leertijd, als de aard van de instructie en het prestatieniveau kunnen variëren. Buss & Poley (1976, p. 195) noemen dit de ideale situatie, waarin elk individu in de gelegenheid gesteld wordt zijn potentiële mogelijkheden te maximaliseren. Een dergelijk onderwijsarrangement dient zeer adaptief te zijn met betrekking tot de individuele verschillen tussen de leerlingen. Een combinatie van de strategie van beheersingsleren en resultaten van aptitude-treatment-interactie (ATI-)onderzoek kan wellicht aanknopingspunten bieden voor de ontwikkeling van dergelijke arrangementen. In hoofdstuk 8 van deze studie komen wij hier nog op terug.

2.4. De Keller-plan-benadering

Ongeveer tegelijkertijd met de ontwikkeling van de strategie van beheersingsleren door Bloom ontwierp de Amerikaan Fred S. Keller een vergelijkbare strategie, die bekend geworden is onder de naam 'Personalized System of Instruction' (PSI) en

ook wel het 'Keller-plan' wordt genoemd. Block (1974, p.18) beschouwt het Keller-plan als een vorm van beheersingsleren ('the second major approach to mastery learning ...'). In een 'invited address' voor de American Psychological Association in 1967, met de wat uitdagende titel 'Good-bye, teacher ...' (Keller, 1968) wordt een opsomming gegeven van de kenmerken van het PSI:

1. individualisering naar tempo: de studenten bepalen zelf hoeveel tijd zij aan het doorlopen van een studietaak willen besteden;
2. vaste criteria voor beheersing van de leerstof: de student mag pas dan beginnen aan een nieuwe leertaak, als de voorafgaande leertaak volledig beheerst wordt;
3. het gebruik van hoorcolleges, demonstraties e.d. als motivatiemiddel in plaats van als primaire bron van informatie-overdracht;
4. sterke nadruk op de geschreven taal in de communicatie tussen docent en student;
5. het gebruik van 'proctors' (hulpleerkrachten) die het mogelijk maken herhaald te toetsen, feedback te geven, extra uitleg te geven etc.

Uit het bovengenoemde derde kenmerk van het PSI blijkt dat de leerkracht als zodanig niet overbodig is, maar slechts dat zijn rol als leerkracht een andere is dan die in meer traditionele onderwijsvormen. In plaats van de primaire informatie-overdrager wordt de leerkracht nu 'engineer or contingency manager of all students' learning' (Block, 1974, p.18-19).

In het Nederlandse onderwijs heeft het Keller-plan vooral invloed gehad op de ontwikkeling van individuele studietoelagen (ISS) in het universitair onderwijs. Alhoewel er duidelijke verschillen zijn aan te wijzen tussen de Nederlandse ISS en de Amerikaanse (zie bijvoorbeeld Van Rookhuijzen et al., 1976, p. 15-21), toch kan gesteld worden dat de in de zeventiger jaren ontwikkelde ISS in het universitaire onderwijs in hoge mate gebaseerd zijn op de uitgangspunten van het Keller-plan (zie bijvoorbeeld Braak, 1974; Plomp, 1974; Verreck, 1973). Het Keller-plan is bij vele verschillende (leer)vakken toegepast, bij diverse vormen van (hoger) onderwijs en in vele landen. Keller (1974) en Keller & Sherman (1974) geven gedetailleerde overzichten van de diverse toepassingen, alsmede van de verschillende aanpassingen die de cursusconstructeurs her en der gemaakt hebben. Voor Nederland geven Van Rookhuijzen et al., (1976) en Plomp & Pilot (1979) soortgelijke overzichten.

2.4.1. Overeenkomsten en verschillen met de beheersings- lerenstrategie

In het universitair onderwijs wordt relatief veel nadruk gelegd op de zelfwerkzaamheid van de student. Dit uitgangspunt heeft ertoe geleid dat bij de vormgeving van het beheersingsleren in individuele studiesystemen sterke nadruk is gelegd op individualisering naar tijd. Bloom daarentegen heeft zijn strategie wat meer uitgewerkt voor het klassikale, groeps-gewijze onderwijs. De overeenkomsten tussen de benaderingen van Bloom en Keller zijn tamelijk groot. Zelfs zodanig, dat het Keller-plan als een bepaalde vorm van mastery-learning beschouwd kan worden. Block (1974, p.20-24) somt een aantal overeenkomsten en verschillen op. Deze zullen hieronder punts-gewijs worden samengevat, en beknopt worden toegelicht.

Overeenkomsten

1. Beide strategieën gaan uit van de veronderstelling dat veel meer studenten een hoog prestatieniveau kunnen behalen dan in traditionele onderwijsvormen gebruikelijk is.
2. Beide benaderingen beginnen ook met het van tevoren specificeren van doelstellingen die elke student in feite moet kunnen bereiken.
3. De wijze waarop het onderwijsarrangement vorm moet krijgen is ook dezelfde in beide strategieën. De werkwijze is als volgt: de cursus wordt opgedeeld in een aantal kleinere leertaken, waarbij elke leertaak slechts bedoeld is de student enkele van de cursusdoelstellingen te laten bereiken. Bovendien moet eerst de ene leertaak met succes voltooid zijn, voordat aan een volgende begonnen mag worden.
Elke leertaak tenslotte bestaat uit twee gedeeltes. Het eerste deel bevat de echte instructie-elementen. Aan de hand van deze leerstof poogt de student voor de eerste maal de doelstellingen te bereiken. Het tweede deel is de zogenaamde feedback- en correctie-component. De functie van deze component is het bewaken van de effectiviteit van de oorspronkelijke leertaak en wanneer blijkt dat deze effectiviteit voor een student te wensen overlaat, dan worden correctieve procedures toegepast.
4. Beide benaderingen gaan er ook van uit dat de diverse studenten ook summatief beoordeeld moeten worden. Deze beoordeling moet gebaseerd worden op wat de student met betrekking tot de cursusdoelstellingen wel of niet geleerd heeft. De individuele beoordeling moet niet afhankelijk zijn van de leerprestaties van medestudenten (absoluut vs. relatief meten (cf. Warries, 1970, 1982)).

Verschillen

1. Zowel bij de Bloom- als bij de Keller-benadering wordt ervan uitgegaan dat de student alle leertaken moet beheersen. Bloom echter baseert de eindbeoordeling (summatief) van een student op een toets die de algemene doelstellingen van de cursus dekt. 'Mastery' wordt dan operationeel gedefinieerd als het behalen van een score gelijk aan of hoger dan een gespecificeerd criterium. Voor Keller echter geldt dat beheersen der delen (de leertaken) gelijk is aan het beheersen van het geheel (de cursus).
2. De omvang van de leertaak is bij Bloom gemiddeld ongeveer twee maal zo groot als bij de Keller-benadering.
3. In de Bloom-strategie wordt veelal gebruik gemaakt van groepsgewijze instructie voor het eerste deel van de instructiecomponent, terwijl in het Keller-plan deze eerste component meestal ook geïndividualiseerd is.
4. In het Keller-plan wordt in de eerste instructie-component meestal alleen maar door de studenten gelezen. Bij de strategie van beheersingsleren zijn ook andere werkvormen mogelijk.
5. In de Bloom-strategie bepaalt de leerkracht (meestal) de instructietijd; in de Keller-strategie doet de student dit.
6. De formatieve toetsen in de Bloom-strategie zijn gedetailleerder van aard dan in de Keller-strategie. Eerstgenoemde toetsen zijn criterium-georiënteerd: elke doelstelling wordt door één of meer items gemeten. De in het Keller-plan gebruikte toetsen zijn algemener van aard en bestaan vaak uit een selectie van items uit een item-verzameling. Niet elke doelstelling behoeft dan aan bod te komen.
7. Het beheersingscriterium bij Bloom ligt altijd lager dan 100%. Bij Keller is dit gelijk aan 100%.
8. Ten aanzien van de correctieve procedures kan in de Bloom-benadering iets meer gedetailleerd worden dan in de Keller-planbenadering. Dit hangt vanzelfsprekend samen met het in punt 6 genoemde onderscheid tussen de diverse formatieve toetsen in beide benaderingen. Een en ander heeft tot gevolg dat in het Keller-plan studenten vaak de gehele leertaak opnieuw moeten bestuderen, terwijl in de Bloom-strategie meestal slechts onderdelen van de leertaak opnieuw gedaan moeten worden. Bovendien is de aard van de correctiva in de beheersingslerenstrategie meestal anders dan in de oorspronkelijke instructiecomponent. In het Keller-plan zijn ze vaak gelijksoortig of identiek. Tierney (1973) heeft beide typen correctiva nader onderzocht. In de volgende paragraaf zullen we onder meer daar aandacht aan

besteden.

Zoals uit bovenstaande lijst van overeenkomsten en verschillen blijkt, zijn de twee benaderingen van origine niet identiek. In de loop der tijd zijn er echter zoveel modificaties in de toepassingen van beide strategieën ontstaan, dan het in de praktijk niet goed mogelijk is een en ander strikt van elkaar te scheiden en te onderscheiden. Block & Burns (1977) maken in hun bespreking van het vele empirisch onderzoek met betrekking tot beide strategieën, dan ook geen strikt onderscheid meer. In het vervolg zullen wij deze werkwijze eveneens volgen. Onder beheersingsleren verstaan wij dus zowel de Bloom-, als de Keller-strategie.

2.5. Resultaten van empirisch onderzoek betreffende de strategie van beheersingsleren

Verschillende auteurs hebben uitgebreide overzichten van onderzoeksresultaten betreffende de effecten van de beheersingsleren-strategie gepubliceerd, o.a. Block (1971, 1974) en Block & Burns (1976). Ook de onderzoeksresultaten betreffende het Personalized System of Instruction zijn door verscheidene auteurs samengevat, onder andere Kulik, Kulik & Carmichael (1974) en Ryan (1974). Met name het artikel van Block & Burns (1976), waarop ook Poortvliet (1979) zich baseert, bevat de resultaten van een zeer groot aantal onderzoeken zowel betreffende het mastery-learning paradigma, als betreffende het PSI-systeem. In het onderstaande laten wij beknopt de conclusies uit het overzichtsartikel van Block & Burns (1976) volgen, waarbij steeds gesproken wordt over beheersingsleren met inbegrip van het Keller-plan.

Uit het meeste onderzoek (o.c., blz. 14-25) blijkt dat onderwijsarrangementen, gestructureerd volgens de principes van de beheersingsleren-strategie, tot betere leerresultaten bij de leerlingen leiden dan meer traditionele vormen van onderwijs. Het systeem 'werkt' dus. De behaalde resultaten zijn echter geringer dan de opstellers der strategie in het algemeen beweren.

Twijfelachtig is voorts of het geleerde ook werkelijk beklijft (Dit geldt overigens, in gelijke mate voor andere didactische methoden). De resultaten van retentiemetingen zijn in dit opzicht niet altijd bemoedigend. Zo vond in Nederland De Bruyne (1976) dat beheersingsleren in een aantal gevallen wel en in een aantal gevallen niet werkte. In een onderzoek naar toepassing van de beheersingslerenstrategie in het statistiek-onderwijs op universitair niveau constateert hij in een tweede retentie-onderzoek dat de studenten '... nog maar bar weinig

kennis en vaardigheden overhadden van de leerstof die zij een jaar daarvoor nog 'spelenderwijs', beheersten. Men moet zich dan toch afvragen wat er schort aan het onderwijs, dat daar op de wat langere termijn zo weinig van blijft hangen' (o.c., p. 169).

Binnen de strategie voor beheersingsleren (incl. de Keller-plan-benadering) wordt onder meer geclaimd dat deze strategie leidt tot positieve affectieve resultaten (Bloom, 1974). In het door Block & Burns samengevatte onderzoek (Block & Burns, 1977, p.26-30) blijkt dat de empirische data tot nu wijzen in de voorspelde richting, alhoewel zij een zogenaamd Hawthorne-effect niet uitsluiten. De instrumenten gebruikt om affectieve resultaten te meten variëren van gestandaardiseerde attitude-schalen tot instrumenten die specifiek voor het betrokken onderzoeksproject geconstrueerd zijn. Soms wordt onder affectief verstaan de houding van de leerling ten opzichte van zijn eigen persoon als student/leerling. Ook werd de affectieve variabele geoperationaliseerd als het aantal leerlingen dat zich voor een bepaalde cursus terugtrok. Dit laatste was met name het geval voor de cursussen opgezet volgens de PSI-benadering.

In hun analyse van de tijdsfactor in verschillende mastery-learning onderzoeken komen Block & Burns (1977, p. 30-33) tot de conclusie dat de langzame studenten sneller gaan studeren, zodat de verschillen in benodigde studietijd tussen de 'slechte' en 'goede' studenten, kleiner worden. In het begin van de beheersingslerenstrategie hebben de zwakkere studenten relatief veel tijd nodig in vergelijking met de betere studenten. Later verdwijnen deze grote verschillen. De beide auteurs geven overigens wel toe dat de gegevens waarop zij deze conclusies baseren minder empirisch onderbouwd zijn dan wenselijk is. De trend is echter wel overtuigend: de strategie van beheersingsleren heeft een nivellerend effect op individuele verschillen in studietijd.

In onderzoek waarbij nagegaan werd of de beheersingslerenstrategie differentiële invloed had op studenten die qua aanvangskenmerken van elkaar verschilden, bleek dat de strategie het effect van dergelijke verschillen niet elimineerde. Soms echter werd wel de invloed van dergelijke verschillen op de uiteindelijke leerresultaten geminimaliseerd (zie bijv. Burrows & Okey, 1975). Uit nader onderzoek (bijv. Anderson, 1976,; Carlson & Minke, 1975; Semb, 1974) bleek dat met name de eis tot beheersing (tot het vereist criterium) van de verschillende leertaken, alsmede het in aansluiting daarop ook daadwerkelijk beheersen van de leerstofinhoud der diverse leertaken, grote invloed had op de leerprestaties der studenten. De oorzaak van deze invloed ligt naar alle

waarschijnlijkheid in de kwaliteit en kwantiteit van de door de student bestede leertijd. Het is dus niet uitsluitend de hoeveelheid leertijd die de student ter beschikking krijgt, het gaat er ook wel degelijk om wat er in die tijd gebeurt (zie onder andere Block & Burns, 1977). Tevens blijkt dat hoe hoger de criteria in de diverse leertaken zijn, hoe groter ook de door de student aan de leertaak bestede studietijd is. En ook: hoe groter de werkelijk bestede studietijd ('time-on-task') is.

Uit onderzoek naar de voorwaarden waaraan voldaan moet worden opdat de invoering van de strategie van beheersingsleren in het onderwijs soepel verloopt, blijkt dat leerkrachten geen totaal nieuw gedragsrepertoire met betrekking tot deze werkvorm behoeven te leren. Het gaat slechts om de toevoeging van een paar nieuwe vaardigheden (bijv. het afnemen van formatieve toetsen en het toepassen van correctieve procedures) aan een reeds bestaand patroon van vaardigheden (Block & Burns, 1977, p. 36-38).

Nuy en Timmer (1979, p. 118-199) stellen echter dat in de vereiste kennis, vaardigheden en attitudes van de docent een hele verschuiving optreedt als hij overgaat van een klassikale naar een meer geïndividualiseerde aanpak.

2.6. Enige kanttekeningen bij de strategie van beheersingsleren

De strategie van beheersingsleren is, zoals uit de voorafgaande paragrafen blijkt, een veelbelovende benadering voor wat betreft de vormgeving van onderwijs-arrangementen. De strategie is beproefd bij een veelheid van (school)vakken en in een grote variatie van onderwijstypen, variërend van de basisschool tot het universitaire onderwijs. Naar de mening van Block & Burns (1977, p.39) biedt het onderzoek met betrekking tot de strategie van beheersingsleren aanknopingspunten voor de ontwikkeling van nieuwe causale theorieën aangaande het leren op school. Tot voor enige jaren waren dergelijke theorieën voornamelijk gebaseerd op leertheoretische noties (zie bijv. Hilgard & Bower, 1975), waarbij de vertaling naar de alledaagse schoolpraktijk vaak minder gemakkelijk verliep dan wellicht wenselijk zou zijn. Met name het gebrek aan ecologische validiteit van dergelijke theorieën is hieraan debet. Overigens moet opgemerkt worden dat ook leerpsychologische theorieën ontwikkeld zijn, waarbij de vertaling van theorie naar praktijk rechtstreeks en met minder problemen geschiedt. Wij doelen hierbij met name op de Sovjetpsychologie, waarbij de bestudering van leerprocessen in

sterke mate plaatsvond vanuit het gezichtspunt van het onderwijzen (zie bijv. Van Parreren & Carpay, 1972, 1980). Ook in de Amerikaanse leerpsychologie van het laatste decennium (bijv. Ausubel, Bruner, Gagné) werd meer nadruk gelegd op een geïntegreerde bestudering van leren en onderwijzen. Het 'mastery-learning' model biedt aanknopingspunten voor een - noodzakelijkerwijze prescriptieve - instructietheorie. Het model richt de aandacht op bepaalde groepen van variabelen, zoals bijvoorbeeld feedback en correctie. Tevens maakt het model duidelijk dat niet alleen de hoeveelheid voor de student beschikbare leertijd van belang is voor de uiteindelijke leerresultaten, maar ook de *kwaliteit* van de in die leertijd ontplooiende leeractiviteiten. Alleen meer leertijd ter beschikking stellen leidt niet zonder meer tot hogere leerprestaties.

Een voor het onderwijs erg belangrijk facet van de strategie van beheersingsleren is de, wat men zou kunnen noemen, *ideologische* grondslag van het model. Deze bestaat hierin dat men er vanuit gaat dat diverse variabelen betreffende het schoolse leren beïnvloedbaar en veranderbaar zijn. Het is met andere woorden geen natuurwet dat slechts weinig leerlingen de onderwijsdoelstellingen kunnen bereiken. Met name dit aspect van de veranderbaarheid van variabelen speelt een rol bij het model van Bloom betreffende het leren op school dat in hoofdstuk 3 besproken wordt. In dit model is gepoogd een theoretische fundering te verschaffen voor het schoolse leren op basis van empirisch onderzoek vooral betreffende het beheersingsleren.

Naast alle positieve geluiden betreffende de mogelijkheden van de strategie van beheersingsleren zijn echter ook meer kritische geluiden te beluisteren. Buss & Poley (1976, p.195-197) spreken zelfs over de 'theoretical bankruptcy of mastery learning'. Deze auteurs menen dat m.n. de notie dat tengevolge van toepassing van de strategie individuele verschillen tussen leerlingen verdwijnen, onjuist is. Voorts stellen zij dat tijd als variabele niet bijster interessant is, daar dit vanuit psychologisch gezichtspunt weinig verklaringswaarde heeft. Ontwikkelingspsychologen beschouwen tijd veeleer als een afhankelijke in plaats van als een onafhankelijke variabele. Het is een variabele die nuttig kan zijn om te beschrijven, maar niet om (leeftijdsafhankelijke) veranderingen te verklaren. Zij menen dat het beschouwen van individuele verschillen in aanleg ('aptitude') in termen van tijd benodigd om het leerstof-beheersingscriterium te bereiken, de onderzoeker ervan weerhoudt te zoeken naar verklaringen en onderliggende processen en variabelen die het differentieel gedrag kunnen verklaren. Voor Bloom is dat, naar hun

opvatting, slechts één variabele: individuele verschillen in benodigde leertijd. Een dergelijk reductionisme achten zij te extreem.

Tegenover dit standpunt kan echter gesteld worden dat het streven naar wetenschappelijke verklaringen met gebruikmaking van zo weinig mogelijk variabelen (begrippen) uit een oogpunt van zuinigheid ('law of parsimony') beslist niet afkeurenswaardig is. Anderzijds zal op den duur de verklaring van individuele verschillen niet beperkt kunnen blijven tot de variabele tijd, daar dit te weinig aanknopingspunten biedt voor nadere theoretische uitwerking. In zijn onderzoek betreffende de toepassing van de beheersingslerenstrategie verwijst De Bruyne (1976, p. 167) naar een symposium van de AERA Special Interest Group 'Research in Social studies Education'. Centraal stond daar de vraag of de strategie van beheersingsleren toepasbaar was in het onderwijs in de sociale wetenschappen. Opvallend was met name de constatering dat we wat beheersingsleren aangaat, nog maar helemaal aan het begin staan. Overheersend was een zeker scepticisme en terughoudendheid betreffende de mogelijkheden van het beheersingsleren (Rice, 1976). Anderzijds wordt ook door De Bruyne (o.c., p.168) gesteld, dat zijn onderzoek in ieder geval heeft aangetoond dat beheersingsleren succes kan hebben. Warries (1979, p. 200) noemt dit het existentiebewijs: beheersingsleren blijkt te bestaan. Theoretische uitbouw en empirische fundering blijven echter noodzakelijk. Het in hoofdstuk 3 van deze studie beschreven model van het schoolse leren is in dit verband een poging tot theoretische uitbouw. Het in deze studie beschreven onderzoek vormt een poging tot nadere empirische fundering van dit model.

EEN MODEL VOOR HET LEREN OP SCHOOL

3.1. *Het model*

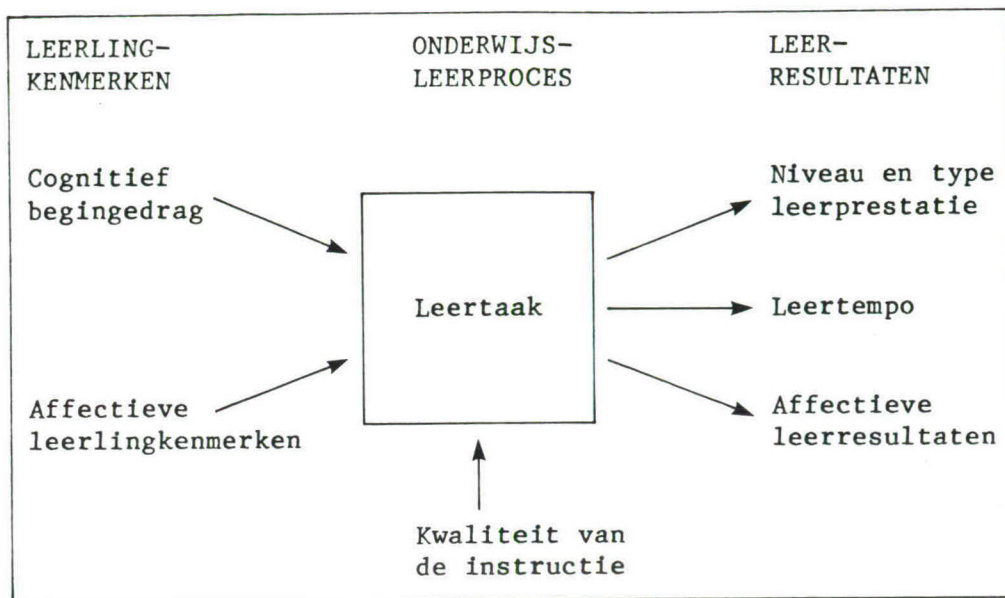
Het in het voorgaande hoofdstuk beschreven onderwijsleermodel van Carroll is een typisch onderwijskundig model in die zin, dat het nauwelijks verwant is aan psychologische theorieën die zich met het leren van individuen onder gecontroleerde omstandigheden bezighouden. Het model gaat uit van de onderwijsleersituaties zoals die in de dagelijkse onderwijspraktijk worden aangetroffen. Ook het model voor het leren op school van Bloom (1976), waarop het in dit proefschrift beschreven onderzoek betrekking heeft, is gebaseerd op het Carroll-model.

Bloom (1976, p. 4) stelt dat uit het Carroll-model blijkt dat als leerlingen met betrekking tot hun aanleg voor een bepaald leerstofonderdeel normaal verdeeld zijn, en zij bovendien allen op dezelfde wijze onderwezen worden, dan zullen hun leerprestaties gemeten na voltooiing van het leerstofonderdeel ook normaal verdeeld zijn. De relatie (correlatie) tussen aanleg ('aptitude') vóór het onderwijsleerproces en leerprestatie gemeten na het onderwijsleerproces zal relatief hoog zijn (ongeveer .70). Omgekeerd zullen als leerlingen normaal verdeeld zijn met betrekking tot aanleg, maar de kwaliteit van het onderwijsleerproces en de beschikbare leertijd worden aangepast aan de kenmerken van de individuele leerling, bijna alle leerlingen het leerstofonderdeel volledig beheersen.

De correlatie tussen aanleg ('aptitude') gemeten vóór afgaand aan het onderwijs en leerprestaties gemeten na afloop van het onderwijs zal dan vrijwel nihil zijn.

In het model van Bloom wordt gepostuleerd dat dit leren een functie is van drie variabelen: het cognitieve begingedrag van de leerling, zijn affectieve aanvangskenmerken en de kwaliteit van het onderwijsleerproces. Variaties in deze - onderling afhankelijke - variabelen bepalen de aard van de leerresultaten, te weten aard en niveau van de schoolvorderingen, de leersnelheid en de affectieve leerlingkenmerken in samenhang met de leertaak. In figuur 3.1 zijn in de vorm van een input-output model de verschillende variabelen schematisch weergegeven.

Als zowel de leerlingkenmerken als het onderwijsleerproces gunstig zijn, dan zullen de leerresultaten ook gunstig zijn en zullen de metingen daarvan relatief weinig variantie vertonen. Als er een behoorlijke spreiding is tussen de leerlingen met betrekking tot hun aanvangskenmerken, en de kwaliteit van het



Figuur 3.1. Bloom's model voor het leren op school

onderwijsleerproces is niet optimaal voor de verschillende leerlingen, dan zal er een zeer grote variantie in de leerresultaten ontstaan.

Het model is volgens Bloom (1976, p. 12) bedoeld om de interactie tussen een individuele leerling, het onderwijsleerproces, de leertaak en de uiteindelijke leerresultaten te verklaren. Tevens is het model bedoeld als verklaringsmodel voor de *verschillen* in de uiteindelijke leerresultaten bij verschillende leerlingen. Het is in deze laatste zin dat het model gebruikt wordt bij de in dit proefschrift beschreven experimenten.

3.2. De leerlingkenmerken

3.2.1. Het cognitief begingedrag

Een leerling begint een onderwijsleersituatie niet als een 'tabula rasa'. Integendeel, de in het verleden opgedane leerervaringen, kennis en vaardigheden bepalen gedeeltelijk de interactie van de leerling met de leertaak en daarmee de leerresultaten van die interactie. Het cognitief begingedrag is nu het geheel van dergelijke leerervaringen benodigd om met goed gevolg een bepaalde leertaak te kunnen voltooien.

Vanzelfsprekend is het zo dat de ene leertaak een groter beroep doet op voorafgaande kennis en vaardigheden dan een andere. Hiërarchisch geordende leertaken, zoals bijvoorbeeld in het wiskunde-onderwijs, zijn in hoge mate gebaseerd op deze onderlinge afhankelijkheid. Ook echter bij minder sequentieel, of hiërarchisch geordende leerstofonderdelen, zoals bijvoorbeeld bij het leren van een tweede taal, gaat men er steeds vanuit dat de leerling beschikt over enige voorafgaande (relevante) leerervaringen. Van belang is niet alleen dat de leerling vroeger relevante leerervaringen heeft opgedaan, maar ook dat hij deze op het gewenste moment tot zijn beschikking heeft. Hij moet ze zich ten minste kunnen herinneren en in de nieuwe leersituatie kunnen gebruiken.

Het cognitief begingedrag is in het model van Bloom predictief met betrekking tot de leerresultaten na afloop van een onderwijsleerproces. Op basis van de analyse van een groot aantal empirische onderzoeken komt Bloom tot de conclusie dat veel leerlingen op school niet beoordeeld worden op hetgeen zij daar leren, maar veeleer beoordeeld worden op grond van hun voorafgaande leerervaringen. Met andere woorden, zij worden in feite beoordeeld op grond van hun kennis en vaardigheden opgedaan in het voorafgaand onderwijs. Aan de hand van longitudinaal onderzoek, waarin een voortoets zowel de effecten van het cognitief begingedrag, als die van de affectieve leerlingkenmerken, alsmede die van hun gezamenlijk effect meet, berekent Bloom de correlatie met een natoets. De range van gerapporteerde correlatiecoëfficiënten loopt van .68 tot .86. Met andere woorden ongeveer tweederde van de variantie in de leerprestaties aan het einde van een cursus wordt enige jaren vóórdat de cursus aanvangt bepaald.

Teneinde de overlap tussen de meting van de leerprestaties en de effecten van de affectieve beginkenmerken te reduceren, is ook nagegaan wat de relatie is tussen aanlegtestscores ('aptitude') en leerprestaties in daarop aansluitende vakken. De aangeboden leerstof vormde het begin van een nieuwe serie leertaken, zodat geen specifiek beroep op bepaalde voorkennis gedaan wordt. De gebruikte aanlegtests bevatten over het algemeen een soort indicator van voorafgaande leerervaringen die relevant zijn voor de te leren vakken of leerstof. Leerlingen met een hoge aanlegtestscore zullen over het algemeen beter presteren op specifieke leerstofonderdelen dan leerlingen met een lagere aanlegtestscore. Met andere woorden de aanlegtest is een indicator voor het cognitief begingedrag en als zodanig predicerend voor latere leerprestaties. De correlatie tussen aanlegtestscores en leerprestaties in aanvangscursussen met betrekking tot een bepaald leerstof-

domein bedraagt gemiddeld .62. De correlatie tussen aanleg en een eerste cursus in een tweede taal bedraagt .60. Bij selectie van een subtest die specifiek voorafgaande leer-gedragingen meet, bedraagt de correlatie .52 voor het leren van een tweede taal.

Ondanks de beperktheid van de aanlegtest, ook nu weer blijkt dat leerprestaties bij nieuwe leerstof gedeeltelijk bepaald worden door leerling-kenmerken die bestonden vóórdat dit onderwijsleerproces begon.

Naast het meer specifieke begingedrag kunnen ook meer algemene aspecten onderscheiden worden. Dit zogenaamde

gegeneraliseerde cognitieve begingedrag omvat variabelen als leesvaardigheid, algemene intelligentie, e.d. Deze aspecten vormen als het ware een gemeenschappelijk begingedrag voor vele schoolvakken voor diverse leerjaren en schooltypen. Uit de International Studies of Evaluation blijkt dat voor vele vakken leesvaardigheid een belangrijk aspect van het begingedrag is (Thorndike, 1973). De mediane correlatie tussen leesvaardigheid en diverse schoolvakken is in de onderbouw van het voortgezet onderwijs .70 en in de bovenbouw .56.

De correlatie tussen schoolvakken valt vrijwel terug tot nul als de factor leesvaardigheid uitgepartialiseerd wordt.

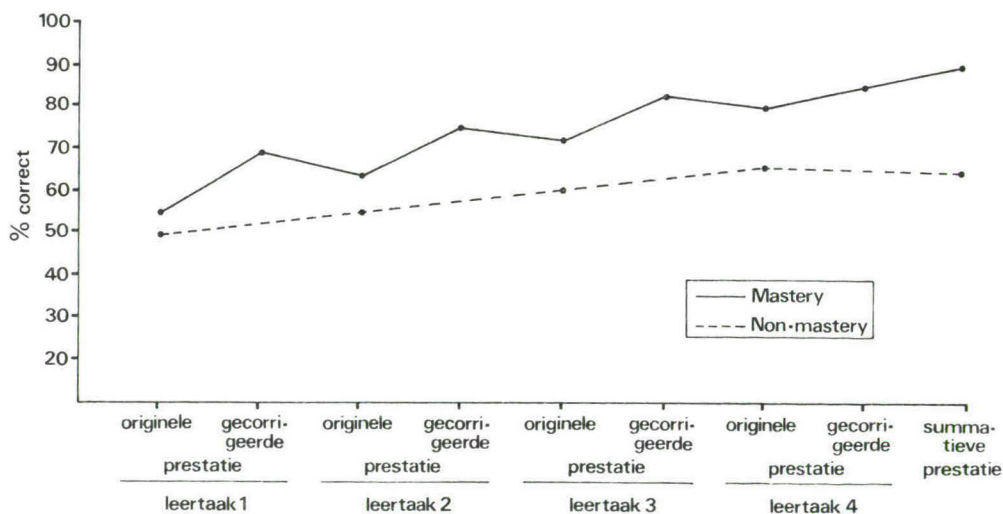
Ook algemene intelligentie voegt weinig toe aan de predictie van schoolvorderingen in vergelijking met het cognitief begingedrag. Algemene intelligentie verklaart ongeveer 25% van de variantie, terwijl het cognitief begingedrag meer dan 50% verklaart (Bloom, 1976, p. 52). Belangrijk is in dit verband ook dat intelligentie (en dit geldt ook voor 'aptitude') nauwelijks door de leerkracht te beïnvloeden is. Als variabele bij het optimaliseren van onderwijsleerprocessen speelt intelligentie derhalve een ondergeschikte rol. Het cognitieve begingedrag daarentegen is in hoge mate beïnvloedbaar, en dus veranderbaar. Immers het omvat kennis van specifieke leerstof, bepaalde vaardigheden e.d. Indien op een of meer van deze aspecten lacunes bestaan, dan kunnen deze middels feedback en correctieve hulp gebracht worden op een zodanig niveau dat de volgende leertaak met succes voltooid kan worden. Het cognitief begingedrag stelt de leerkracht in staat het leer-verleden van de leerling te relateren aan de nieuwe leer-situatie. Intelligentietests bieden in dit opzicht weinig houvast.

Bloom (1976) bespreekt ook een groot aantal experimenten waarbij het eindgedrag van de ene leertaak het begingedrag vormt van de volgende. Het in de hoofdstukken 7 en 8 van dit proefschrift beschreven onderzoek komt qua opzet sterk overeen met deze experimenten. Bloom spreekt in dit verband van 'micro-level studies'. Steeds wordt daarbij een 'mastery-

learning'-benadering vergeleken met een meer traditionele onderwijsleersituatie. Voorafgaand aan elke leertaak wordt het cognitief begingedrag gemeten. Aan het einde van elke leertaak wordt een (formatieve) toets afgenomen. Voor de mastery-conditie volgden vervolgens nog enige correctieve procedures (zie par. 3.3.), waarna wederom een formatieve paralleltoets wordt afgenomen. Op deze wijze is er dan voor gezorgd dat het cognitieve begingedrag voor de volgende leertaak adequaat is in de mastery- en minder adequaat in de niet-mastery-conditie. De mediane correlatie tussen leerprestaties op de ene en die op de volgende leertaak bedraagt .67. Deze waarde komt overeen met eerder gevonden correlaties.

De mediane correlatie tussen de leerprestaties na correctie in de mastery-learning-conditie en de leerprestaties in de volgende leertaak vóór correctie bedraagt .32. Derhalve beduidend lager, maar nog niet gelijk aan nul zoals theoretisch verwacht.

In figuur 3.2. wordt grafisch weergegeven hoe het prestatieverloop over 4 leertaken in een mastery- en non-mastery-conditie is in een experiment bij het leren van een tweede taal (Binor, 1974, geciteerd bij Bloom, 1976, p. 59).



Figuur 3.2. Veranderingen in leerprestaties over leertaken

Duidelijk blijkt dat de leerprestaties op elke leertaak veranderbaar zijn.

Hypothetisch kan gesteld worden dat als alle leerprestaties

systematisch voor alle leerlingen (indien nodig) gecorrigeerd worden, zodat allen het vereiste criteriumniveau behalen met betrekking tot het cognitieve begingedrag, dan zal het effect van het cognitieve begingedrag op opeenvolgende leertaken tot nul gereduceerd worden.

3.2.2. *Affectieve leerlingkenmerken*

Bloom verstaat onder de affectieve leerlingkenmerken de attitude en de interesse van leerlingen om aan een bepaalde reeks van leertaken te beginnen. Het is een bekend gegeven dat er tussen leerlingen nogal wat verschillen bestaan met betrekking tot deze kenmerken, zelfs vóórdat zij enig onderwijs met betrekking tot een bepaald leerstofdomein genoten hebben. Het definiëren van wat men exact onder affectieve leerlingkenmerken verstaat, is nogal moeilijk. In het Carroll-model wordt de motivatie voor een leertaak omschreven als de tijd die de leerling aan die taak wil besteden. Waarom de leerling dat wil, wordt echter niet duidelijk. De term motivatie is wat te algemeen voor het begrip affectieve leerlingkenmerken, terwijl bovendien de rechtstreekse relatie met de leertaak onduidelijk is. Desondanks worden onder affectieve leerlingkenmerken ook allerlei motivationele factoren begrepen.

Bloom (1976, p. 104) schat dat affectieve leerlingkenmerken ongeveer 25% van de variantie in cognitieve leerprestaties verklaren. Voor het onderwijs in een tweede taal kent Jacobovits (1970) in het totale schema van variabelen die leereffect bepalen, aan de motivatie een waarde van 33% toe. Ongetwijfeld zit hierin ook een deel van het effect van het cognitieve begingedrag besloten.

Het effect van affectieve leerlingkenmerken op de leerprestaties zal in het algemeen pas na langere tijd merkbaar zijn. Bijvoorbeeld op de leerprestaties aan het einde van een schooljaar. Anderzijds is het niet uitgesloten dat de invloed ook merkbaar is na een korte tijdsperiode, bijvoorbeeld na afloop van een aantal (samenhangende) lessen.

De affectieve leerlingkenmerken kunnen verdeeld worden in een drietal aspecten, te weten affect met betrekking tot een (school) vak, attitudes ten aanzien van (de) school en attitudes ten aanzien van de eigen persoon of ego. Het eerstgenoemde aspect is het minst, het laatstgenoemde is het meest algemeen. In de begin-schooljaren kunnen deze aspecten nog wel onderscheiden worden, maar geleidelijk aan raken zij steeds meer geïntegreerd.

Uit de International Studies of Evaluation blijkt dat de mediane correlatie tussen de attitude ten opzichte van een schoolvak en de leerprestaties in dat vak variëren van .32 tot .38 in het achtste leerjaar ('grade 8') en van .30 tot .52 in het twaalfde leerjaar ('grade 12'). Een uitzondering moet gemaakt worden voor het vak letterkunde waar de correlatie aanmerkelijk lager is (.23, resp. .15). Carroll (1975) rapporteert voor Nederland correlaties tussen leerprestaties in het schoolvak Frans en de attitude ten opzichte van dit vak van resp. .26 in het achtste leerjaar en .33 in het twaalfde leerjaar. In het algemeen blijkt de attitude ten opzichte van een schoolvak ongeveer 10% van de variantie in leerprestaties voor dat vak te verklaren (Bloom, 1976, p. 85).

Eveneens op basis van de International Studies of Evaluation komt Bloom (o.c., p. 92) tot de conclusie dat de correlatie tussen attitude ten aanzien van de school en leerprestaties in de hogere leerjaren gemiddeld .45 bedraagt. Attitude ten opzichte van school verklaart derhalve ongeveer 20% van de variantie in leerprestaties.

De attitude ten aanzien van de eigen persoon ('academic self concept') blijkt in staat ongeveer 25% van de leerprestaties na de basisschool te verklaren (o.c., p. 95).

De tot nu toe gepresenteerde gegevens zijn voornamelijk gebaseerd op, wat Bloom noemt, 'macro-studies'. Deze strekken zich uit over een relatief lange tijdsperiode.

In het kader van de mastery-learning experimenten (zie par. 3.2.1.) zijn Bloom c.s. ook de relatie nagegaan tussen leerprestaties en de attitude met betrekking tot het (school) vak in relatief kort durende tijdsperiodes. Bij de berekening der correlatie-coëfficiënten zijn de mastery- en niet-mastery-leerlingen gecombineerd.

De mediane correlatie tussen attitude aan het begin van een leertaak en leerprestaties aan het einde van een leertaak bedraagt .30 (o.c., p. 98). In het algemeen blijkt uit de diverse 'micro-studies' dat de attitude ten aanzien van het (school)vak en de leerprestaties op gelijke wijze correleren als in de macro-studies gebleken is.

Als algemene conclusie kan gesteld worden, dat de relatie tussen affectieve leerlingkenmerken en leerprestaties ongeveer .50 bedraagt. Affectieve leerlingkenmerken blijken ongeveer 25% van de variantie in cognitieve leerprestaties te kunnen verklaren (o.c., p. 104).

3.3. *Kwaliteit van de instructie*

In het algemeen hebben de leerlingkenmerken aan het begin van het onderwijsleerproces meer invloed op de latere leerprestaties dan de kwaliteit van de instructie. Het zgn. effectiviteitsonderzoek naar de invloed van leerkrachtvariabelen op de leerprestaties van leerlingen heeft niet zo erg veel opgeleverd (Rosenshine & Furst, 1973). Slechts een gering percentage van de variantie in leerlingprestaties kon verklaard worden.

In het Carroll-model wordt de kwaliteit van de instructie aangeduid in termen van de mate waarin de presentatie, uitleg en ordening van taakelementen het optimum voor een bepaalde leerling benadert. Bloom geeft een iets andere inhoud aan het betreffende begrip. De kwaliteit van de instructie heeft te maken met de aanwijzingen ('cues') die de leerling krijgt, met de participatie van de leerling in de leeractiviteiten, met de beloning ('reinforcement') die de leerling krijgt m.b.t. de leertaak en tenslotte met een terugkoppelings- en correctiesysteem.

De 'cues' zijn de aanwijzingen die de leerlingen krijgen over wat en hoe er geleerd moet worden. De aanwijzingen worden soms gegeven aan het begin van een onderwijsleerproces, bv. in de vorm van 'advance organizers' (Ausubel 1968), of in de vorm van doelstellingenlijsten, soms echter ook in de loop van een onderwijsleerproces. Het onderzoek m.b.t. dit aspect heeft veel minder te maken met (persoons)kenmerken van leerkrachten, wat centraal stond in het onderzoek naar 'teacher-effectiveness', maar veel meer met datgene wat leerkrachten daadwerkelijk doen.

'Reinforcement', positieve of negatieve bevestiging, is een begrip dat in vele theorieën en modellen m.b.t. het leren een plaats heeft. Het grootste gedeelte van het onderzoek heeft betrekking op de relatie tussen leerprestaties en typen 'reinforcement', de frequentie van beloning en de omvang en aard van de beloning voor verschillende leerlingen in een groep.

Met het begrip 'participatie' geeft Bloom aan dat de leerling iets moet doen met de diverse 'cues' daar er anders geen leren plaatsvindt. Op basis van de resultaten van een twintigtal onderzoeken komt Bloom (1976, p. 123) tot de conclusie dat 20% van de variantie in de leerprestaties van individuele leerlingen verklaart kan worden op grond van verschillen in participatie in het onderwijsleerproces in de klas.

Bij het terugkoppelingssysteem ('feedback') noemt Bloom in één adem de zgn. correctieve procedures.

3.3.1. *Feedback*

De term feedback wordt in het model van Bloom in algemene zin gebruikt. In navolging van Kulhavy (1977) gebruiken wij de term feedback eveneens in algemene zin ter aanduiding van de talrijke procedures die gebruikt worden om een leerling duidelijk te maken dat een door hem gegeven antwoord in een leertaak goed of fout is. Op welke wijze de feedback verschaft wordt, blijft in deze omschrijving buiten beschouwing.

We kunnen een drietal factoren onderscheiden die van invloed zijn op het feedbackproces. Allereerst de *voorkennis* (cf. Kulhavy, 1977) van de leerling. (In de terminologie van Bloom zouden we nu kunnen spreken over het cognitief begin-gedrag.) Feedback geeft niet alleen informatie aan de leerling omtrent de juistheid of onjuistheid van zijn antwoord, maar het verschaft hem indirect ook informatie over de mate waarin hij erin is geslaagd de in de leertaak gepresenteerde leerstof te verwerken. Als de voorkennis van de student met betrekking tot de aangeboden leertaak onvoldoende is, dan biedt informatie over het al dan niet juist zijn van het gegeven antwoord geen aanknopingspunten voor een betere leerprestatie. Een tweede factor van invloed op het feedbackproces wordt gevormd door de *mogelijkheden tot correctie*. Indien in het onderwijsleerproces mogelijkheden geboden worden om gemaakte fouten te verbeteren dan verhoogt dit de effectiviteit van de terugkoppeling. (In par. 3.3.2. zal met name ingegaan worden op deze correctieve procedures.) Indien een leerling te horen krijgt dat een door hem gegeven antwoord fout is, dan zal hij trachten deze fout te herstellen. Het is echter zeer de vraag of de leerling daar in slaagt, zonder aanvullende informatie betreffende de aard van de fout, betreffende de verkeerde redenering die tot het foute antwoord geleid heeft en dergelijke. In Nederland heeft met name De Klerk (1974, 1977) gewezen op de noodzaak om leerlingen niet alleen informatie te verschaffen over het al dan niet correct zijn van hun antwoord, maar ze ook te informeren over het proces dat tot het juiste antwoord zou moeten leiden. Anderzijds blijkt uit sommige onderzoeken (bijv. Merrill, 1970; Merrill et al., 1970) dat meer specifieke informatie als feedback niet automatisch leidt tot hogere leerprestaties. Wel bleek overigens dat de leerefficiëntie in termen van benodigde leertijd werd verhoogd.

Een derde en laatste factor tenslotte wordt gevormd door de *hoeveelheid feedback*. Hoe meer informatie de leerlingen over hun leerproces krijgen, hoe beter ze in staat zijn dit proces te optimaliseren (Kulhavy, 1977). Elke medaille heeft zijn keerzijde, dus ook deze. Een te frequent gebruik van

feedback kan zelfs nadelige effecten hebben. Indien een leertaak overladen wordt met mogelijkheden om feedback te geven, dan wordt daardoor de aandacht van de eigenlijke leertaak teveel afgeleid (zie bijv. Talyzina, 1973). In hoeverre feedback ook daadwerkelijk effectief is (dat wil zeggen leidt tot hogere leerprestaties), hangt af van de wijze waarop de terugkoppeling vorm wordt gegeven.

Er moet sprake zijn van echte terugkoppeling, en wel zodanig dat de leerling niet zonder meer de beschikking moet hebben over het juiste antwoord op een gestelde vraag. Als het antwoord te dicht en te duidelijk bij de betreffende vraag staat, dan denkt de leerling onvoldoende na. Kulhavy (1977) duidt dit aan met de term 'presearch availability'. Onderzoek (bijv. Anderson & Faust, 1967) toont aan dat bij een hoge presearch availability de leerlingen eenvoudigweg het antwoord overnemen en de instructie overslaan. Als de availability echter laag is, dan moeten de leerlingen de leerstof grondig bestuderen voordat ze de gestelde vragen kunnen beantwoorden (cf. Anderson, Kulhavy & Andre, 1981; 1972).

In de literatuur wordt met betrekking tot de wijze van terugkoppeling ook aandacht besteed aan de grootte van het tijdsverloop tussen het moment dat het antwoord is gegeven en de feedback plaatsvindt. De zogenaamde 'delayed feedback' (dat wil zeggen terugkoppeling enige tijd nadat het antwoord is gegeven) blijkt vaak tot betere resultaten te leiden dan onmiddellijke feedback, (cf. Sassenrath & Young, 1969; Sturges, 1969). Kulhavy en Anderson (1972) baseren onder meer hierop hun mening dat feedback niet de rol vervult van beloning (reinforcement) zoals in de klassieke behavioristische opvattingen veelal wordt gesteld. Een belangrijk aspect, waarop reeds eerder werd gewezen, heeft betrekking op het feit dat de terugkoppeling moet kunnen leiden tot correctie van de gemaakte fouten. Dit betekent meestal dat de gegeven feedback informatief moet zijn. Immers de leerlingen moeten op basis van de verstrekte informatie hun leerproces kunnen bijsturen. Informatieve feedback blijkt dan ook meestal beter te werken dan de simpele goed/fout mededeling (cf. Gilman, 1969). Aan de hand van de verstrekte informatie (bijv. verwijzingen naar leerstofonderdelen, aanvullende uitleg) kan de leerling dan de gemaakte fouten herstellen.

Tenslotte moet met betrekking tot de terugkoppelings-technologie nog gewezen worden op het onderscheid actief/passief. Bij actieve feedback moet de leerling daadwerkelijk wat met de aangereikte informatie doen, zoals bijv. tekst-onderdelen nog eens bestuderen, of een alternatief schema doornemen. Bij meer passieve feedback krijgt de leerling te horen wat het goede/foute antwoord en - bij een fout antwoord

- wat het correcte antwoord zou moeten zijn. Vaak spreekt men dan van 'kennis van correcte resultaten' (KCR). Dit is een terugkoppelingsprocedure die in het traditionele onderwijs toch wel tamelijk vaak gehanteerd wordt. In een onderzoek van Tait, Hartley & Anderson (1973) werden een drietal feedback-procedures vergeleken, t.w. geen feedback, passieve feedback en actieve feedback. Uit analyse van de onderzoeksresultaten bleek wel dat feedback tot hogere leerprestaties leidde dan geen feedback, maar geen significant verschil kon aangetoond worden tussen actieve en passieve feedback.

3.3.2. *Correctieve procedures*

Hoe goed het onderwijsleerproces ook is opgebouwd, steeds zal blijken dat niet alle leerlingen de leerstof tot het vereiste criterium beheersen. Op zich is dit natuurlijk niet verwonderlijk daar de instructie meestal bedoeld is voor een groep leerlingen. En wat voor de ene leerling geschikt is, behoeft dat voor een andere leerling nog niet te zijn. In een optimaal onderwijsleerproces behoren dus ook voorzieningen getroffen te zijn om leerlingen die de leerstof nog niet volledig beheersen alsnog tot het vereiste beheersingsniveau te brengen. Dergelijke voorzieningen worden *correctieve procedures* ('correctives') genoemd.

De meeste onderzoeksliteratuur ten aanzien van correctieve procedures is te vinden in het domein van de literatuur met betrekking tot het beheersingsleren. Het meeste onderzoek op het terrein van het beheersingsleren gaat na of er verschil bestaat in leerprestaties tussen groepen die wel en die niet volgens de strategie van het beheersingsleren zijn onderwezen. Het belangrijkste verschil tussen de groepen werd dan meestal gevormd door het wel of niet gebruiken van correctieve procedures (Block, 1971, 1974).

De correctieve procedures omvatten suggesties aan de leerling welke onderdelen van de leerstof nog eens opnieuw bestudeerd moeten worden, ofwel er wordt nieuw instructiemateriaal betreffende de onvolledig beheerste leerstof aangereikt, zoals bijvoorbeeld geprogrammeerde instructies, additionele leerboeken, schema's, e.d. In elk geval krijgt de leerling extra tijd en instructie om alsnog het vereiste beheersingsniveau te bereiken.

Block & Anderson (1975) stellen dat correctieve procedures dezelfde leerstof moeten onderwijzen als in de oorspronkelijke onderwijsleersituatie, maar dat het op een andere wijze moet gebeuren dan in de oorspronkelijke instructie. Zij refereren niet aan empirisch onderzoek op basis waarvan deze stelling

verdedigd kan worden. Uitgaande van het Carroll-model kan men ook stellen dat leerlingen die nog niet voldoen aan het gestelde criterium nog wat meer tijd nodig hebben.

De door Block & Anderson gegeven opsomming van correctieve procedures, zoals alternatieve leerboeken, audiovisuele hulpmiddelen, 'token economy system', geprogrammeerde instructie, tutor-systemen, etc., doet nogal willekeurig aan. Het lijkt er een beetje op, alsof voor de leerlingen die in eerste instantie een onvoldoende behalen, allerlei leuke (leer-) activiteiten ontwikkeld worden. Binnen een traditionele onderwijssetting zijn verschillende der correctieve procedures ook wat moeilijker in te plannen. Het verdient veel meer de voorkeur om correctieve procedures te kiezen die nauw aansluiten bij de oorspronkelijke instructie. Inpassing in reguliere onderwijssystemen is dan eenvoudiger en kansrijker. In het in de eerste alinea van deze subparagraaf genoemde onderzoek, worden groepen vergeleken waarbij wel en geen sprake is van correctieve procedures. Op basis van een achttal onderzoeken rapporteert Bloom (1976, p. 129 e.v.) dat de mediane correlatie tussen kwaliteit van de instructie (wel/geen correctieve procedure) en toetsscores op de laatste formatieve toets (vóór correctie) .47 bedraagt, en die voor de summatieve toetsscore .53.

Duidelijk wordt dat het gebruik van correctieve procedures in de beheersingsleren-groep en het ontbreken van deze procedure in de andere (controle-)groep leidt tot een toenemende differentiatie tussen deze twee groepen. Heel duidelijk komt deze trend naar voren in het onderzoek van Binor (1974) gewijd aan het leren van een tweede taal. De correlaties tussen wel/geen correctieve procedures en leerprestaties op een eerste, een tweede en een derde formatieve toets, alsmede op een summatieve toets bedragen resp. .24, .45, .55 en .65.

In het algemeen blijkt (Bloom, 1976, p. 135) dat kwaliteit van de instructie ongeveer 20-25% van de variantie in leerlingprestaties kan verklaren.

3.4. Leerresultaten

Bij de bespreking van de output-kant van het model van Bloom besteden wij alleen aandacht aan de affectieve leerresultaten en aan het leertempo. Enerzijds omdat de relatie van de cognitieve leerresultaten met de leerlingkenmerken en kwaliteit van de instructie reeds meermalen aan bod is geweest en anderzijds omdat in par. 3.5. aan een en ander nog nader aandacht wordt besteed in een poging de diverse bevindingen in het model te integreren.

3.4.1. *Affectieve leerresultaten*

Bij de bespreking van de affectieve leerlingkenmerken (par. 3.2.2.) is vnl. de relatie nagegaan met (cognitieve) leerprestaties. Met andere woorden centraal stond de vraag naar de voorspellende waarde van affectieve leerlingkenmerken voor cognitieve leerresultaten. Anderzijds kan ook nagegaan worden welke de invloed is van het onderwijsleerproces, in samenhang met initiële leerlingkenmerken, op de attitude van de leerlingen na afloop van het onderwijsleerproces. Dit is de vraag naar de affectieve leerresultaten.

Bloom (1976, p. 140) stelt dat de affectieve consequenties zich relatief langzaam ontwikkelen. Dit betekent dat het in het algemeen moeilijk is om bij relatief kortdurende experimenten veranderingen in attitudes en houdingen van leerlingen te constateren en deze toe te schrijven aan het onderwijsleerproces.

Block & Burns (1976) analyseerden een negental onderzoeken waarbij de mastery-learning-strategie was toegepast en eveneens een negental onderzoeken waarbij de 'Personalized System of Instruction'-benadering was gevolgd. In het algemeen bleek dat beide strategieën bij de leerlingen leidden tot een positieve houding t.o.v. het vak, tot een verhoging van het zelfvertrouwen en tot een positievere attitude t.o.v. het onderwijsleerproces. Ook bleek echter dat naast positieve, ook negatieve gevolgen te onderkennen waren. De strategie voor beheersingsleren leidt tot een significant hogere mate van toetsangst ('test anxiety') dan in meer traditionele instructiestrategieën. Bij de PSI-benadering valt op dat er in vergelijking met de controlegroepen nogal wat leerlingen zich uit de cursus terugtrekken. Dit laatste wordt mogelijkwerwijs veroorzaakt door het feit dat leerlingen zelf hun tijd mogen indelen. Daardoor beginnen ze vaak te laat en trekken zich na enige tijd terug in plaats van een lager cijfer te accepteren. Bloom (1976) draagt weinig empirische evidentie aan ter ondersteuning van zijn stelling dat er een (causale) relatie bestaat tussen de aard en vorm van het onderwijsleerproces en affectieve leerresultaten. Zijn redenering komt er in feite op neer dat negatieve leerervaringen ertoe leiden dat leerlingen minder zelfvertrouwen krijgen en derhalve een minder positieve attitude opbouwen m.b.t. het betreffende (school)vak en de school. Voorkomt men dergelijke negatieve leerervaringen, dan zal dat leiden tot een positieve attitude, hetgeen vervolgens weer een gunstige invloed heeft op de (cognitieve) leerprestaties etc.

3.4.2. *Leertempo*

De belangstelling voor de factor tijd als een belangrijke variabele in het onderwijsleerproces is gewekt door de publicatie van Carroll (1962, 1963). Carroll heeft er met name op gewezen dat het vooral van belang is hoeveel tijd leerlingen daadwerkelijk aan de leertaak besteden. En niet, hoeveel tijd zij er aan mogen besteden. Laatstgenoemde tijdsfactor ('roostertijd') heeft slechts een zeer geringe verklarende waarde voor verschillen in leerprestaties tussen leerlingen in dezelfde klas of op dezelfde school. In het Carroll-model wordt een causale relatie verondersteld tussen de werkelijk bestede tijd en leerprestaties. Bloom (1980) heeft er met name op gewezen dat de werkelijk bestede tijd ('time-on-task') door het onderwijs-leerproces in positieve of negatieve zin veranderd kan worden, en dat deze wijziging directe gevolgen heeft voor de leerprestaties. Met andere woorden als we de leertijd kunnen wijzigen, kunnen we daarmee de leerprestaties beïnvloeden.

Anderson (1980) beschrijft de geschiedenis van het onderzoek met betrekking tot de variabele tijd in onderwijsleerprocessen.

Hij onderscheidt een vijftal benaderingen.

1. *de ecologisch-psychologische benadering*

Aan het onderzoek binnen deze aanpak zijn de namen van Gump (1967) en Kounin (1970) verbonden. Het onderzoek van Gump toonde aan dat 'time-on-task' in het reguliere onderwijs bij gehele schoolklassen geringer was dan bij kleinere door leerkrachten begeleide groepen. Hetzelfde gold voor activiteiten waarbij de leerlingen zelf hun tijd mochten indelen in vergelijking met leerlingen die dat niet mochten.

Gump concentreerde zich voornamelijk op activiteiten in de klas en de tijdsindeling. Kounin trachtte bepaalde leerkrachtgedragingen te ontdekken die geassocieerd waren met hoge niveaus van 'time-on-task'. Met name leerkrachtgedragingen die zorgen voor een zekere continuïteit in het onderwijsleerproces waren geassocieerd met een hoge mate van 'time-on-task' bij de leerlingen.

2. *de 'mastery-learning' benadering*

Het is voor de hand liggend, dat binnen de strategie voor beheersingsleren, gebaseerd op het Carroll-model, aandacht wordt besteed aan het tijdsaspect. In tegenstelling tot de eerstgenoemde benadering is het onderzoek hier veel globaler van karakter. Het is dan ook veel moeilijker om aan te geven welke kenmerken van het onderwijsleerproces gerelateerd zijn aan leertijd. Het duidelijkst

onderscheiden onderwijsleersituaties volgens de strategie van het beheersingsleren zich van die van volgens andere benaderingen door de aanwezigheid van 'feedback en correctieve procedures' en van het duidelijk overbrengen van verwachtingen ten aanzien van leerprestaties en leerprocessen van de leerlingen.

Opmerkelijke resultaten van het onderzoek zijn dat leerlingen binnen de strategie voor beheersingsleren steeds minder tijd nodig hebben om bepaalde prestatiestandaards te behalen, en bovendien steeds homogener worden qua omvang van het geleerde en qua behoefte aan leertijd (Arlin, 1973; Block, 1970; Anderson, 1976).

3. *de 'survival skills'-benadering*

Het onderzoek concentreert zich op het onderwijs in de zogenaamde 'survival skills' aan leerlingen van de basisschool. Cobb (1972) heeft middels een correlatieel onderzoek vier van dergelijke vaardigheden vastgesteld, te weten (1) opletten, (2) werken aan activiteiten die producten opleveren, (3) gedrag vertonen waaruit de leerkracht kan afleiden dat de leerling aan bepaalde activiteiten wil meedoen en (4) meegaandheid: doen wat de leerkracht zegt. Onderzoek toonde aan dat het versterken ('reinforcement') van deze vaardigheden tot resultaat heeft dat de 'time-on-task' verhoogd wordt en daarmee ook de leerprestaties. Reinforcement is hier dus de kritische instructievariabele.

4. *de benadering van de 'Beginning teacher evaluation study' (BTES)*

De eerste fase van de BTES omvatte het opstellen van het algehele design. De tweede en derde fase betroffen de uitvoering van het onderzoek zelf. De tweede fase werd uitgevoerd door Frederick McDonald bij de Educational Testing Service (ETS) te Princeton. De derde fase was gesitueerd aan het Far West Laboratory for Educational Research and Development.

Er werden drie typen leertijd onderscheiden, te weten (1) toegestane tijd ('allocated time'): tijd die beschikbaar is voor een bepaald onderwerp, of cursus, (2) bestede tijd ('time-on-task') en (3) academische leertijd: tijd besteed aan leertaken met een hoog succesniveau.

Uit het onderzoek kwamen twee kritische instructievariabelen naar voren. De eerste variabele wordt aangeduid met de term 'monitoring'. Onderwijsleersituaties waarin de leerkracht de informatie- en materiaalstroom kanaliseert, waarin hij de activiteiten van de leerlingen stuurt, waarin hij middels de juiste terugkoppeling de antwoorden van de leerlingen stuurt, gaan samen met een hoog niveau van

'time-on-task'.

De tweede instructievariabele kan aangeduid worden met de term 'taakafstemming'. Die onderwijsleersituaties waarin de leertaak qua moeilijkheidsgraad afgestemd is op de capaciteiten van de leerlingen, zijn geassocieerd met hoge niveaus van 'time-on-task'. Er is sprake van afstemming als veel leerlingen de criteriumscore behalen.

Samenvattingen van de onderzoekresultaten zijn te vinden in Denham & Lieberman (1980) en in een themanummer van 'The Journal of Classroom Interaction' (Vol. 17, no. 1, 1981), met name in de bijdragen van Fisher et al. (1981), Rosenshine (1981) en Cooley & Mao (1981). Een goed overzicht wordt ook geboden door Veenman (1980).

5. de *'Follow-through evaluation'*-benadering

Het werk van Jane Stallings (1975) is toonaangevend voor deze aanpak. De toegestane tijdsperiodes worden opgedeeld in tijdsperiodes voor specifieke activiteiten. Zo is bijvoorbeeld tijd besteed aan het werken met een leerboek gerelateerd aan leerprestaties.

Klassen waarin leerlingen in groepen van acht met een leerkracht werkten, presteerden beter dan klassen waarin leerlingen in kleinere groepen of alleen werkten. Alle vormen van academische feedback (prijzen, positieve en negatieve correctieve terugkoppeling) waren positief gerelateerd aan leerprestaties.

Deze benadering integreert in feite de eerder genoemde benaderingen. Het streeft naar een synthese.

Opvallend in de vijf genoemde benaderingen is, dat steeds naar voren komt dat effectieve vormgeving van het onderwijsleerproces er toe kan leiden dat de tijd die de leerling werkelijk aan leren besteed, toeneemt. Als een van de kenmerken van een effectief instructiesysteem wordt steeds een terugkoppelings- en correctiesysteem genoemd.

Ook het onderzoek van Anderson (1976) toonde aan dat de door de leerling benodigde tijd om leerstof tot op criteriumniveau te beheersen, beïnvloed kan worden door een effectieve instructieprocedure, zoals bv. de methode van het beheersingsleren.

Bloom (1974, 1980) vergelijkt de 'time-on-task' van leerlingen in een onderwijsleerproces opgezet volgens de principes van het beheersingsleren met die van leerlingen in een traditioneel onderwijsleerproces. Hij constateert dat de variantie in 'time-on-task' gedurende de beginfase van het onderwijsleerproces driemaal zo hoog is in de conditie gestructureerd volgens het beheersingsleren dan in de andere conditie. Tegen het einde van het onderwijsleerproces is het

verschil aanmerkelijk gereduceerd. De verhouding is dan ongeveer 1,5 : 1.

Er is ook een duidelijke relatie tussen de werkelijk aan het onderwijsleerproces bestede tijd en de winst in leerprestaties. Op basis van een analyse van een vijftiental onderzoeken vond Bloom (1976) een correlatie van .40 tussen bestede tijd en leerprestaties als de eenheid van analyse de leerling was en van .52 als de analyse-eenheid de klas was.

3.5. Het model geïntegreerd

Op basis van de in de voorgaande paragrafen gepresenteerde theoretische verwachtingen en empirische evidentie kunnen hypothesen geformuleerd worden ter toetsing van het door Bloom geconstrueerde model. In het volgende hoofdstuk worden deze hypothesen in detail vermeld.

Verwacht kan worden dat de leerprestaties van leerlingen onder optimale instructie-condities hoger zullen zijn dan die van leerlingen onder niet optimale condities. Voorts zal de variantie in leerprestaties van de eerstgenoemden lager zijn dan die der laatstgenoemden. Wat wel of niet optimale instructie-condities zijn, wordt bepaald door de variabele 'quality of instruction'. Hierbij konden een viertal aspecten onderscheiden worden (par. 3.3.), waarvan terugkoppeling en correctieve procedures de belangrijkste waren. Als de leerlingen bovendien beschikken over het juiste cognitieve begingedrag dan zullen de leerprestaties nog hoger zijn. Tenslotte zullen bij positieve affectieve aanvangskenmerken de cognitieve leerresultaten nog beter en de variantie in leerprestaties nog minder zijn.

Overigens kan men zich afvragen of de verwachtingen op basis van het model ook stand houden nadat gecorrigeerd is voor verschillen in leertijd. In dit verband stelt Bloom dat zowel de hoeveelheid benodigde leertijd als de variantie in leertijd afneemt naarmate het leerproces vordert.

De stijging van de leerprestaties, het afnemen van de variantie in leerprestaties en leertijd, etc. wordt binnen het model verklaard met behulp van drie (onafhankelijke) variabelen: het cognitief begingedrag, de affectieve beginkenmerken en de kwaliteit van het onderwijsleerproces. Deze variabelen worden geacht verklarende waarde te hebben voor de geconstateerde verschillen tussen (groepen) leerlingen in reguliere onderwijsleersituaties, waarin opeenvolgende leertaken doorlopen worden.

Het cognitief begingedrag verklaart ongeveer 50% van de variantie in leerprestatie. De affectieve beginkenmerken,

bestaande uit een tamelijk breed spectrum omvattende belangstelling voor en attitude ten opzichte van de leerstof, de school en het onderwijs in het algemeen, alsmede de houding van de leerling ten opzichte van zichzelf, worden geacht plm. 25% van de variantie in leerprestaties te verklaren.

Bloom beschouwt de aan een leertaak voorafgaande leer-geschiedenis van een leerling als bestaande uit een combinatie van affectieve en cognitieve kenmerken. Deze combinatie verklaart gezamenlijk ongeveer 65% van de variantie in leerprestaties tussen de leerlingen (Bloom, 1976, p. 171). De derde variabele, kwaliteit van de instructie verklaart op zich plm. 25% van de optredende variantie, terwijl de drie variabelen tezamen maar liefst 90% van de variantie in leerprestaties kunnen verklaren.

Teneinde het model als geheel empirisch te onderbouwen analyseert Bloom een twaalfstal onderzoeken waarin steeds twee groepen met elkaar vergeleken worden. De groepen krijgen les volgens gelijke procedures, behalve dat de ene conditie extra tijd en hulp krijgt op basis van de resultaten van afgenomen (formatieve) toetsen. Beide groepen worden tenslotte vergeleken op een zelfde summatieve toets. Steeds blijkt de groep die extra tijd en hulp ontvangen heeft gemiddeld beter te scoren. (Extra tijd en hulp in aanmerking genomen zeker geen onverwacht resultaat.) Tevens blijkt echter ook dat de mediane variantie voor de experimentele groep slechts 60% is van die voor de controle-groep. Ook Block & Burns (1977) komen tot een dergelijke conclusie.

3.6. Enige kritische kanttekeningen

Het door Bloom gepropageerde, en in dit hoofdstuk betrekkelijk beknopt beschreven model, is aantrekkelijk door zijn eenvoud. Toch ligt hierin ook een punt van zwakte. Immers de diverse begrippen in het model zijn nogal ruim omschreven. Tot het begrip 'affectieve beginkenmerken' kan een waar arsenaal van gevoelens, houdingen en gedragingen gerekend worden. Hetzelfde, alhoewel in afnemende mate, geldt voor de begrippen 'cognitief begingedrag' en 'kwaliteit van de instructie'. Wat dit laatste begrip betreft lijkt het aspect 'feedback en correctiva' het meest cruciaal te zijn. De omschrijving van het begrip is echter veel ruimer. De wat ruime begrips-definiëring heeft er ook toe geleid dat de operationalisaties van de begrippen-zoals-bedoeld in variabelen-zoals-bepaald velerlei vormen hebben aangenomen. Het is dan ook niet altijd zonder meer evident dat aangetoonde effecten van deze empirische variabelen op een of meer afhankelijke variabelen

kunnen worden beschouwd als empirische evidenties voor het model. Feit is ook dat integrale empirische toetsing van het model tot nu toe niet heeft plaatsgevonden. De wat fluïde begrippen zullen hieraan zeker mede debet zijn.

De door Bloom aangedragen empirische evidentie, gebaseerd op vele tientallen kleine en grotere onderzoeken, verdient bewondering. Zelden ziet men dat diegenen die een theoretisch model presenteren zich ook zoveel moeite getroosten om dat model in ruime mate empirisch te onderbouwen. Indachtig het voorafgaande kan echter ook gesteld worden, dat de ruime begripsomschrijvingen ook alle mogelijkheden bieden tot het inpassen van een veelheid aan onderzoeken.

Imponerend zijn zonder meer de percentages verklaarde variantie voor de verschillen in leerprestaties die Bloom claimt voor de verschillende onafhankelijke variabelen (of begrippen!) in zijn model. In dit verband moet echter gekeken worden naar de door Bloom gevolgde procedure bij het infereren van deze percentages. Bloom zet de correlaties tussen aanvangskenmerken (cognitief en/of affectief), kwaliteit van de instructie, en leerprestaties gevonden in een groot aantal onderzoeken op een rij. Vervolgens berekent hij van al deze correlaties de mediane correlatie, kwadrateert deze en komt aldus (na vermenigvuldiging met 100%) tot het door elke variabele verklaarde percentage variantie.

Door op deze wijze gebruik te maken van de mediane correlatie wordt geen rekening gehouden met de spreiding die er over de diverse onderzoeken bestaat. Nadere beschouwing van de correlatie-coëfficiënten in de diverse onderzoeken leert dat de range soms aanmerkelijk kan zijn. Bij de interpretatie van het percentage verklaarde variantie had op zijn minst met deze range rekening gehouden moeten worden.

De in diverse onderzoeken gevonden correlatie-coëfficiënten vormen de basis voor de percentages verklaarde variantie. Bij deze percentageschattingen gaat Bloom er stilzwijgend vanuit dat de populaties waarnaar gerefereerd wordt, gelijk of vergelijkbaar zijn. Niets is echter minder waar: nu eens hebben de data betrekking op basisschoolleerlingen, dan weer op universiteitsstudenten. Als er gesproken wordt over percentages verklaarde variantie, dan is het noodzakelijk te weten op welke populatie een en ander betrekking heeft. Op dit punt zou grotere nauwkeurigheid verlangd mogen worden (cf. Peterson, 1977).

Ook moet nog gewezen worden op het feit dat Bloom niet of nauwelijks rekening houdt met de tijd die leerlingen nodig hebben om bepaalde leerprestaties te bereiken. Indien de tijd voor alle leerlingen gelijk is, dan zal de ene leerling ongetwijfeld meer en betere prestaties leveren dan de andere

leerling. Met andere woorden de variantie in leerling-prestaties neemt niet af in de mate als door Bloom voorspeld. Te weinig wordt bij de beoordeling van de empirische gegevens met deze factor rekening gehouden. Opvallend is tenslotte ook nog dat Bloom bij de bespreking van de factor 'kwaliteit van de instructie' nauwelijks gebruik maakt van de veelheid van empirische gegevens verzameld in enige decennia (onderwijs-) psychologisch onderzoek, bijvoorbeeld op het terrein van de terugkoppeling.

VRAAGSTELLINGEN

4.1. Inleiding

Sinds het ontstaan van de strategie voor beheersingsleren (Bloom 1968), en dit geldt ook voor het Personalised System of Instruction (Keller, 1968), heeft er nogal wat onderzoek plaatsgevonden naar het functioneren van deze benaderingen. Dit onderzoek is in de loop der tijd langzamerhand van karakter veranderd. Block & Burns (1977) onderscheiden in dit verband vier onderzoekstypen.

Het *Type I-onderzoek* was gericht op de algemene vraagstelling 'Werkt het?'. Nagegaan werd of leerlingen die onderwijs kregen volgens de methode van het beheersingsleren in relatief grotere aantallen de leerstof leerden beheersen dan leerlingen die les kregen volgens meer traditionele onderwijsmethoden. In het *Type II-onderzoek* werd meer gelet op allerlei neveneffecten van de strategie voor beheersingsleren. In het onderzoek ging het onder meer om vragen als: leren de leerlingen leren? Welke tijds-investeringen zijn per leerling benodigd? Etc.

De literatuurgegevens op dit gebied zijn nogal divers van aard en soms tamelijk speculatief. Uit beide typen onderzoek werd in ieder geval duidelijk dat de strategie voor beheersingsleren positieve resultaten voor de leerlingen opleverde. Er ontstond nu een derde research-benadering: het *Type III-onderzoek*. In dit type onderzoek werd nagegaan of de strategie voor beheersingsleren dezelfde effecten heeft op leerlingen met verschillende cognitieve en affectieve beginkenmerken. Voorts werd nagegaan of bepaalde componenten van de strategie grotere effecten hebben op de leerlingen dan andere componenten. In vergelijking met de eerder genoemde onderzoekstypen zijn de onderzoeksdesigns veel complexer van aard. Tenslotte is er dan nog het *Type IV-onderzoek*, dat vrijwel gelijktijdig met het Type III-onderzoek ontstond. Dit onderzoek richt zich voornamelijk op de toepassing en implementatie van de strategie voor beheersingsleren in de praktijk. Een voorbeeld van het Type IV-onderzoek in het Nederlandse onderwijs is bijvoorbeeld Nuy (1981).

Het in deze dissertatie beschreven onderzoek behoort deels tot het Type I- en deels tot het Type III-onderzoek. Het Type I-onderzoek heeft voornamelijk betrekking op de toetsing van het in het voorafgaande hoofdstuk beschreven model van Bloom betreffende het leren op school. Nagegaan wordt of het model klopt voor wat betreft het leren van een tweede taal, door

leerlingen die reeds enige voorkennis betreffende deze taal hebben. Het Type III-aspect van het onderzoek komt naar voren in dat gedeelte waarin de variabele 'kwaliteit van de instructie' gevarieerd wordt.

Met name wordt nagegaan wat de effecten zijn van verschillende soorten 'terugkoppeling en correctieve procedures' op leerprestaties, tijdsbesteding e.d. De Type III-benadering biedt de mogelijkheid om de empirische bevindingen van het onderwijspsychologische onderzoek te integreren in het algemene model. Deze benadering komt ook naar voren, daar waar nagegaan wordt wat de effecten zijn van verschillende correctieve procedures op leerlingen met verschillende cognitieve aanvangskenmerken.

4.2. Vraagstellingen betreffende het model

Binnen het in het voorgaande hoofdstuk besproken model van Bloom wordt door een drietal inputvariabelen, te weten het cognitief begingedrag, de affectieve leerlingkenmerken en de kwaliteit van de instructie plm. 90% van de variantie in leerlingprestaties op een summatieve toets verklaard.

In het kader van het in hoofdstuk 7 besproken experiment wordt nagegaan in hoeverre dit percentage verklaarde variantie onderbouwd wordt door empirische bevindingen.

Niet alleen wordt het percentage verklaarde variantie bepaald met betrekking tot een summatieve toets, maar ook met betrekking tot een retentietoets die enige tijd later wordt afgenomen. Immers uit eerder onderzoek (zie bijv. De Bruyne, 1976) is bekend dat de leerresultaten enige tijd na afname van de summatieve toets nogal eens te wensen overlaten. Voordat het percentage verklaarde variantie bepaald wordt, zal eerst nagegaan worden of het verloop van het leerproces der proefpersonen overeenkomstig het model is. Immers, indien bijv. na afname van een formatieve toets en toepassing van een correctieve procedure geen stijging van de leerprestaties geconstateerd kan worden, dan functioneert een en ander niet zoals door het model beoogd. Het percentage verklaarde variantie der leerprestaties kan dan niet meer vergeleken worden met de door Bloom gerapporteerde en voorspelde percentages.

4.3. Vraagstellingen met betrekking tot de cognitieve leerresultaten

In het derde hoofdstuk zijn bij de bespreking van het model

reeds diverse relaties tussen leeractiviteiten en leerresultaten aan de orde geweest. Bij de formulering van de algemene vraagstellingen in deze paragraaf kunnen we over de aanleiding derhalve kort zijn.

Vraagstelling 1.

Hebben leerlingen die onder kwalitatief betere instructie-procedures onderwijs hebben gehad betere leerresultaten dan leerlingen die onder slechtere condities onderwijs hebben gehad?

Met kwalitatief betere instructie-procedures wordt een onderwijsleerproces bedoeld waarin op enigerlei wijze de component 'terugkoppeling en correctiva' een plaats heeft.

Vraagstelling 2.

Neemt bij leerlingen die onder kwalitatief betere instructie-procedures onderwijs hebben gehad de variantie in leerprestaties af ten opzichte van leerlingen die anderszinds onderwijs hebben gehad?

Het gaat in dit verband vooral om de variantie in leerresultaten op de formatieve toetsen.

Vraagstelling 3.

Neemt bij leerlingen die volgens kwalitatief betere instructieprocedures les hebben gehad het verband tussen aanleg ('aptitude') en leerprestaties af?

Deze vraagstelling gaat uit van de veronderstelling dat de leerprestaties van de leerlingen onder optimale leeromstandigheden minder afhankelijk zullen zijn van de aanleg van de leerlingen. Met name dit aspect is in de literatuur met betrekking tot het beheersingsleren sterk benadrukt. Het is ook een aspect waarop de vormgevers van onderwijsleerprocessen zich moeten richten. (Overigens zijn de mogelijkheden 'natuurlijk' niet onbegrensd!).

Te verwachten is dat het door het cognitief begingedrag verklaarde percentage variantie in leerprestaties lager zal zijn naarmate het onderwijsleerproces meer gestructureerd wordt. Immers de instructie-component zal dan een grotere invloed op de leerresultaten moeten krijgen. Een dergelijke bevinding zou in overeenstemming zijn met de in het mastery-learning-onderzoek aangedragen empirische evidentie betreffende de afnemende correlatie tussen 'aptitude' en leerprestaties.

4.4. Vraagstellingen met betrekking tot de affectieve modelcomponent

Bloom meent dat een optimaal onderwijsleerproces bevorderend werkt op een positieve houding van leerlingen. Met name zou

dit moeten gelden voor de wat minder goede leerlingen. Immers de leerprestaties van de leerlingen worden niet beoordeeld in het licht van de prestaties van andere leerlingen, maar er wordt alleen nagegaan of ze de leerstof beheersen. Waarna, zo nodig, bijsturing volgt. Aldus kan geformuleerd worden:

Vraagstelling 4.

Bevordert een optimaal onderwijsleerproces een positieve houding van de leerling ten opzichte van de school, de leerstof en zichzelf? Is dit effect sterker voor de (cognitief) laag presterende leerlingen in vergelijking met de (cognitief) hoog presterende leerlingen?

4.5. Vraagstellingen met betrekking tot het leertempo=de leertijd

Overeenkomstig het model van Bloom en de opvattingen van Carroll betreffende de leertijd, kan geformuleerd worden:

Vraagstelling 5.

Worden de verschillen tussen de leerlingen in de benodigde tijd om een leertaak te voltooien geringer terwijl zij onderwijs middels een kwalitatief betere procedure volgen in vergelijking met leerlingen die volgens een kwalitatief minder goede procedure les krijgen?

Ook hier wordt weer met 'kwalitatief beter onderwijs' bedoeld onderwijs waarin aandacht besteed wordt aan feedback en correctie. In algemene zin kan de in par. 4.3.2. geformuleerde vraagstelling nader gedifferentieerd worden door rekening te houden met de factor tijd. Het is immers voor de hand liggend dat leerlingen betere prestaties leveren naarmate zij meer leertijd toegewezen krijgen. Het maken van formatieve toetsen en hertoetsen, plus het ondergaan van enige correctieve procedures kost tijd. Bij het vergelijken van leerprestaties tussen groepen leerlingen moet derhalve rekening gehouden worden met deze extra tijd.

Het is juist dit aspect dat veelal in het onderzoek met betrekking tot het beheersingsleren veronachtzaamd wordt. Meestal vindt geen controle plaats voor de verschillen in leertijd tussen de diverse onderzochte condities.

Het is bovendien te verwachten dat de in par. 4.2. en 4.3. bedoelde percentages verklaarde variantie voor leerlingen binnen een beheersingsleren-conditie lager zullen zijn. Met andere woorden de variantie-verklarende kracht van het model zal lager zijn indien rekening gehouden wordt met de (waarschijnlijk) benodigde extra leertijd. Overigens wordt in de literatuur (zie bijv. Bloom 1974) ook gepostuleerd dat aanvankelijk extra leertijd geïnvesteerd moet worden, maar dat

deze na verloop van tijd weer als het ware teruggewonnen wordt, doordat de leerlingen betere leermethoden ontwikkelen, met andere woorden ze leren leren.

4.6. Vraagstellingen met betrekking tot de kwaliteit van de instructie

De component 'kwaliteit van de instructie' kan op velerlei wijzen gevarieerd en tot object van onderzoek gemaakt worden. Wij zullen ons beperken tot het aspect 'feedback en correctieve procedures', aangezien onder meer door Block & Burns (1977) wordt aanbevolen dit aspect nader in het onderzoek met betrekking tot het beheersingsleren te betrekken.

Ten aanzien van de correctieve procedure zal een onderscheid gemaakt worden tussen een zogenaamde Keller- en een Bloom-type. In het eerste geval gaat het dan om totale herhaling van de (oorspronkelijke) leerstof, terwijl het in het tweede geval gaat om partiële herhaling en een andere leerstofpresentatie. Eerder onderzoek (bijv. Tierney, 1973) geeft geen eenduidige evidenties betreffende de kwaliteit van de ene t.o.v. de andere methode. Met betrekking tot de feedback-procedure zal onder meer nagegaan worden of feedback überhaupt tot betere leerprestaties leidt. Tenslotte wordt nagegaan of feedback waarin een actieve houding van de leerling verlangd wordt: hij moet iets met de leerstof doen, tot hogere prestaties leidt dan feedback waarbij met een meer passieve houding volstaan kan worden.

Ook bij deze laatste vergelijking moet in feite gecorrigeerd worden voor (eventuele) verschillen in benodigde leertijd. Immers het actief met de leerstof omgaan kost waarschijnlijk meer tijd dan de wat passievere benadering. De vraagstellingen betreffende de feedback en correctieve procedures zullen dan ook beantwoord worden met de verschillen in leertijd in aanmerking genomen.

Hoofdstuk 5

HET GEBRUIKTE CURSUSMATERIAAL

5.1. Inleiding

In dit hoofdstuk wordt een beschrijving gegeven van de in het kader van dit onderzoek gebruikte cursus. Besloten werd een bestaande cursus te gebruiken en deze - waar nodig -aan te passen aan de door ons onderzoek gestelde eisen. Door deze werkwijze werd gegarandeerd dat een ecologisch valide onderwijsleersituatie geschapen werd, waardoor generalisatie van de experimentele bevindingen naar andere, soortgelijke cursussen eenvoudiger is. Een nadeel van het gebruiken van een nieuwe, zelfontwikkelde cursus is bovendien dat dergelijk materiaal in de regel nauwelijks uitgetoetst en geëvalueerd kan worden.

Gezien het feit dat met name in ons onderzoek aandacht besteed werd aan de door de proefpersonen gebruikte leertijd, moest de cursus wel de mogelijkheid tot individualisering bieden. Gekozen werd dan ook voor een zodanige cursusopzet, waarbij de proefpersonen de diverse onderdelen naar eigen inzicht en in eigen tempo konden bestuderen.

5.2. Inhoud en vormgeving van de cursus

5.2.1. Relevantie

Aan verschillende universiteiten en hogescholen in Nederland wordt onderwijs in vreemde talen gegeven aan niet-vakstudenten. Soms volgen deze studenten dit vreemde talenonderwijs uit louter persoonlijke belangstelling, soms echter is dit onderwijs in het kader van hun hoofdvakstudie verplicht. Zo moeten bijvoorbeeld hoofdvakstudenten 'Internationale economie' aan de Katholieke Hogeschool Tilburg verplicht deelnemen aan een of meer cursussen moderne vreemde talen.

Bij de vormgeving van dergelijke cursussen is een tweetal overwegingen uitgangspunt geweest. De eerste is dat het vreemde talenonderwijs bruikbaar moet zijn voor het uitoefenen van het later beroep. Het tweede uitgangspunt is dat de cursussen moeten aansluiten op de middelbare schoolkennis van degenen die een bepaalde taal in hun eindexamenpakket hebben opgenomen.

In de praktijk van het vreemde talenonderwijs aan niet-vakstudenten aan universiteiten en hogescholen blijkt dat

grammatica-onderricht daarin steeds op de een of andere manier een plaats inneemt. Ook aan de Katholieke Hogeschool Tilburg werd in de praktijk van het onderwijs de behoefte gevoeld aandacht te besteden aan grammatica-onderwijs, in ieder geval voor het Frans, maar later ook voor de talen Duits en Engels. Ook in cursussen die geschreven worden voor volwassenen die al eens een vreemde taal geleerd hebben, maar hun kennis daarvan willen ophalen of uitbreiden, wordt meestal ook aandacht besteed aan grammatica.

5.2.2. Opzet van de cursus

De (ten dele gemodificeerde) cursus waarvan in het onderzoek gebruik is gemaakt, is in eerste instantie door medewerkers van het Talencentrum der K.H.T. opgezet ter ondersteuning van de colleges Frans. De cursus zou echter ook geschikt moeten zijn voor hen die in het talenpracticum door zelfstudie 'hun Frans willen ophalen', en ook ter ondersteuning van eventueel later nog te ontwikkelen colleges. Het moest dan ook mogelijk zijn om vanuit welke cursus dan ook te verwijzen naar deze cursus. Er is naar gestreefd de grammaticacursus, waarover het hier gaat, op te delen in korte eenheden. De keuze van de grammatica-onderwerpen is gebaseerd op foutenanalyses van de resultaten op een vrije spreektoets afgelegd door studenten die het college Frans op de Katholieke Hogeschool volgden in 1974-1975, alsmede op de foutenanalyse uitgevoerd door Knibbeler (1975). In ons onderzoek kon verwacht worden, dat proefpersonen/studenten die de cursus zouden volgen niet bij voorbaat al reeds alles beheersten, maar ook, dat zij bij serieuze studie in staat geacht moesten worden een behoorlijk rendement te halen.

In het Zweedse Cume-project ontstond de indruk dat de cognitieve leertheorie een betere basis biedt voor het grammatica-onderwijs aan volwassenen dan de audio-linguale gewoontevormingstheorie. Ook Ulijn (1975) komt tot deze conclusie. Bij de vervaardiging van de cursus is hiermee rekening gehouden.

Elke onderwijsleereenheid, een op zichzelf staand onderdeel van de cursus, wordt ingeleid door een theoretisch gedeelte gevolgd door een aantal schriftelijke oefeningen, in het algemeen van het invultype. (In de oorspronkelijke cursus zijn ook nog mondelinge oefeningen - structuurdills - opgenomen, die hier weggelaten zijn.) De bedoeling van de schriftelijke oefeningen is de gebruiker in staat te stellen, de regels die hij geleerd heeft, in alle rust toe te passen.

De schriftelijke oefeningen worden gevolgd door sleutels,

Tabel 5.1. Inhoudsopgave lessen en lesonderdelen

Les	Onderwerp	Par.	Onderwerp
1	Het lidwoord	1.1.	Het bepaald lidwoord
		1.2.	Gebruik van het lidwoord voor namen van landen.
		1.3.	Het delend lidwoord
2	Het aanwijzend voornaamwoord	2.1.	Het bijvoeglijk gebruikte aanwijzend voornaamwoord.
		2.2.	Het vervangend gebruikte aanwijzend voornaamwoord.
		2.3.	Het gebruik van 'ce', 'il' en 'cela' bij werkwoorden.
		2.4.	Het gebruik van 'ce' en 'il' bij het werkwoord 'être'.
3	Het bezittelijk voornaamwoord	3.1.	Het 'bijvoeglijk gebruikte' bezittelijk voornaamwoord.
		3.2.	Het 'zelfstandig gebruikte' bezittelijk voornaamwoord.
4	Het betrekkelijk voornaamwoord	4.1.	Het betrekkelijk voornaamwoord als onderwerp of als lijdend voorwerp.
		4.2.	Het betrekkelijk voornaamwoord na een voorzetsel.
5	Het persoonlijke voornaamwoord	5.1.	Het persoonlijk voornaamwoord.
		5.2.	De plaats van het persoonlijk voornaamwoord in de zin.
		5.3.	De onderlinge plaatsing in de zin van twee persoonlijke voornaamwoorden.
6	Hulpwerkwoorden	6.1.	'Avoir' en 'être'.
		6.2.	Andere hulpwerkwoorden
7	De lijdende vorm	-	-

waarbij niet alleen het juiste antwoord gegeven wordt, doch middels een verwijzing ook wordt aangegeven op welk gedeelte van de theorie de betreffende oefenvraag betrekking heeft. De proefpersoon/student kan in zijn tekst de regels vinden die in bepaald item toegepast moeten worden. Op deze wijze is het mogelijk geworden dat elke les individueel en zelfstandig door de student/proefpersoon doorgewerkt kan worden.

De cursus waarmee in het kader van dit onderzoek gewerkt werd, bestond uit een selectie van 7 onderwijsleereenheden uit de oorspronkelijke cursus. Elke les werd opgedeeld in een aantal onderdelen. In tabel 5.1. wordt een overzicht gegeven van de diverse lessen en hun onderdelen/paragrafen.

5.2.3. Doelstellingen

Bij het oorspronkelijk cursusmateriaal ontbraken eenduidig en expliciet geformuleerde doelstellingen. Teneinde voor gebruik in het onderzoek formatieve, dus leerdoelgerichte, toetsen te kunnen construeren was het echter noodzakelijk deze doelstellingen a posteriori te formuleren.

Voor elke onderwijsleereenheid werd een leerstofanalyse uitgevoerd op basis waarvan doelstellingen werden geformuleerd.

De geformuleerde doelstellingen werden vervolgens ter beoordeling voorgelegd aan een medewerker van het Talencentrum der Katholieke Hogeschool Tilburg. Gevraagd werd te beoordelen of de geformuleerde doelstellingen de leerstof volledig representeerden en of de inhoudelijke formulering correct was. Na enige aanvulling en bijstelling werd de representativiteit van de doelstellingen ten opzichte van de leerstof als voldoende beoordeeld.

Op grond van de beoordeling van de inhoudsvaliditeit der doelstellingen werden bij nader inzien ook nog enige lacunes ontdekt in enkele leerstofonderdelen. Door aanvulling en bijstelling konden deze deficiënties opgeheven worden.

5.2.4. Formatieve toetsen

Op basis van de voor elke onderwijsleereenheid geformuleerde doelstellingen werden toetsitems geconstrueerd. Voor elk item werd nauwkeurig aangegeven met welke doelstelling het correspondeerde. Door een vakdeskundige van het Talencentrum der K.H.T. werd beoordeeld of de items enerzijds taalkundig qua vocabulaire en grammaticale regels juist geconstrueerd waren en anderzijds of de items 'pasten' bij de diverse

doelstellingen. Op deze wijze werden garanties geschapen met betrekking tot de inhoudsvaliditeit der items.

Bij elke doelstelling werden meerdere items gemaakt. Vervolgens werden deze items aselekt (per doelstelling) verdeeld over twee toetsen. De eerste toets werd aan het einde van een les afgenomen en diende ter bepaling van het beheersingsniveau. Indien dit niveau niet voldeed aan bepaalde, in het onderzoek gedefinieerde, normen dan moest de student/proefpersonen nog enig aanvullend onderwijs volgen. Na afloop daarvan werd dan de tweede toets afgenomen.

Het was vanzelfsprekend de bedoeling dat de beide toetsen gelijkwaardig waren. Met betrekking tot deze gelijkwaardigheid kunnen we een tweetal aspecten onderscheiden:

- meten de toetsen in leerinhoudelijke zin hetzelfde en
- zijn de toetsen in psychometrische zin als paralleltoetsen te beschouwen.

Aangaande het eerste aspect kan gesteld worden dat de geschapen garanties met betrekking tot de inhoudsvaliditeit der items alsmede door de aselekte toewijzing van de items aan de toetsen het alleszins redelijk is te veronderstellen dat de toetsen in leerinhoudelijke zin vergelijkbaar zijn.

De vraag of de toetsen ook in psychometrische zin als parallel beschouwd kunnen worden, is moeilijker te beantwoorden. Parallelliteit wordt in de klassieke testtheorie (zie bijv. Lord & Novick, 1968, p. 47-48) tamelijk strikt omschreven. Drenth (1975, p. 203 e.v.) formuleert een drietal voorwaarden waaraan voldaan moet worden alvorens meetinstrumenten als parallel beschouwd kunnen worden. De tests moeten (1) dezelfde standaarddeviatie hebben, (2) gelijke correlatie vertonen met een reeks ware scores en (3) de variantie die niet wordt gedetermineerd door de ware scores moet opgevat worden als toevalsvariantie. Van deze drie aannames is alleen de eerste rechtstreeks empirisch aantoonbaar.

De vraag blijft echter of deze definitie van parallelliteit afkomstig uit de klassieke testtheorie ook moet gelden voor de door ons gebruikte formatieve toetsen en her-toetsen. Het is een bekend gegeven (zie onder meer hoofdstuk 6 van dit proefschrift) dat de klassieke testtheorie minder geschikt is voor de psychometrische analyse en beoordeling van zogenaamde leerdoelgebonden of criterium-georiënteerde toetsen. Het gaat veelal om de consistentie van beslissingen die genomen worden op basis van de op een formatieve toets behaalde score.

Wij gaan er nu vanuit dat de formatieve her-toetsen, op identieke wijze geconstrueerd als de formatieve toetsen, parallel zijn met de laatstgenoemde. Een mogelijkheid dit empirisch te onderbouwen zou kunnen zijn de slaag/zak-

beslissingen voor dezelfde leerlingen, genomen op basis van deze twee toetsen, met elkaar te vergelijken.

Bij de constructie van de toetsen werden twee uitgangspunten gehanteerd. De toetsen mochten niet te omvangrijk worden, daar anders de benodigde toetstijd disproportioneel zou worden en elke doelstelling moest om redenen van representativiteit getoetst worden. Per doelstelling in een toets werd volstaan met twee items. De vorm der items kwam overeen met die van de items zoals die in de verschillende lessen opgenomen waren. Het item bestond uit een zin in de Franse taal waarin een woord was weggelaten. Dit woord had een grammaticale functie

Tabel 5.2. Aantal items per les en per lesonderdeel in de formatieve toetsen en her-toetsen

Les	Totaal aantal items	Les- onderdeel*	Aantal items
1. Het lidwoord	30	1.1.	10
		1.2.	10
		1.3.	10
2. Het aanwijzend voornaamwoord	24	2.1.	4
		2.2.	8
		2.3.	4
		2.4.	8
3. Het bezittelijk voornaamwoord	22	3.1.	14
		3.2.	8
4. Het betrekkelijk voornaamwoord	22	4.1.	12
		4.2.	10
5. Het persoonlijk voornaamwoord	78	5.1.	50
		5.2.	12
		5.3.	16
6. Hulpwerkwoorden	20	6.1.	6
		6.2.	14
7. De lijdende vorm	12	-	12

* Zie voor de paragraafinhoud tabel 5.1.

en kon worden ingevuld op grond van de betreffende grammaticale regel. Bij sommige toetsen werd in de instructie een opsomming gegeven van in aanmerking komende in te vullen woorden. In tabel 5.2. wordt een overzicht gegeven van de aantallen items per toets per les en per lesonderdeel. De correctie der toetsen geschiedde objectief en de score werd bepaald door per goed beantwoord item 1 punt toe te kennen en deze punten per onderdeel en per toets te totaliseren.

5.2.5. *Remediële instructie*

Afhankelijk van de score der studenten/proefpersonen op de formatieve toetsen, en afhankelijk van de conditie in het experiment waaraan de proefpersonen toebedeeld waren, werd hen een of andere vorm van remediële instructie verstrekt.

Bij de keuze der correctieve procedures is uitgegaan van een drietal criteria (Weeda 1978):

1. de correctieve procedures dienden zo dicht mogelijk aan te sluiten bij de aard en inhoud van de gebruikte instructieprocedure en het instructiemateriaal;
2. het moest voor de student/proefpersoon inzichtelijk zijn dat bij een onvoldoende initieel instructieresultaat een correctieve procedure een logische aanvulling was;
3. de aard en vorm der correctieve procedures diende zoveel mogelijk aan te sluiten bij de in de literatuur voorhanden zijnde (empirische) evidentie.

Toepassing van het eerste criterium leidde er enerzijds toe dat in sommige condities van de experimenten de gehele oorspronkelijke les, dan wel een of meer onderdelen daarvan, opnieuw bestudeerd moesten worden. Anderzijds leidde toepassing van het eerste criterium ertoe dat een schematisch overzicht gemaakt werd van de in de les, of onderdelen van de les, behandelde leerstof. In het eerste geval sluit de correctieve procedure maximaal aan bij het oorspronkelijke leermateriaal. Verwacht kan echter worden dat door het ontbreken van een nieuwheidsprikkel er een wat meer passieve leerhouding bij de student/proefpersoon ontstaat. Bij de laatstgenoemde correctieprocedure is de inhoud niet, doch de wijze van presentatie, ordening en dergelijke wel nieuw. De verwachting was dan ook dat een dergelijke correctieve procedure een meer actieve leerhouding zou uitlokken.

Het tweede criterium 'inzichtelijkheid voor de student/proefpersoon' werd vervuld door bij de introductie van de lessen aan te geven dat een zeker minimumprestatieniveau behaald moest worden en dat bij falen enige aanvullende instructie noodzakelijk was met betrekking tot het onderwerp waarvoor het

bereikte leerresultaat tekort schoot.

Toepassing van het derde criterium 'aansluiten bij in de literatuur voorhanden zijnde empirische evidentie' tenslotte, stuit op enige problemen. Er is namelijk relatief weinig of geen empirische evidentie met betrekking tot dit aspect voorhanden. Block & Anderson (1975) stellen, dat correctieve procedures '... must teach the same material as does your groupbased plan for the unit, but they must do so in ways that differ from this plan' (p. 33).

De remediële instructie moest met andere woorden niet alleen een aanvulling zijn op de voorafgaande instructie, maar eigenlijk ook een alternatief daarvoor zijn. Wat betreft de door ons gebruikte correctieve procedures voldoet het schema wel, en de herhaling van (onderdelen van) de les niet aan deze eis. Anderzijds wordt ook gesteld (o.c., p. 38) dat bij oudere leerlingen/studenten individuele correctieve procedures beter voldoen dan groepswijze. Alle door ons gebruikte correctieve procedures voldoen hier wel aan.

PSYCHOMETRISCHE ASPECTEN DER FORMATIEVE TOETSEN

In het voorgaande hoofdstuk is aandacht geschonken aan de inhoud en vormgeving van de in ons onderzoek gebruikte cursus. Een belangrijk onderdeel van deze cursus werd gevormd door de leerstofgebonden toetsen. Het gebruik van dergelijke toetsen in moderne curricula en onderwijssystemen neemt de laatste jaren steeds meer toe. Van louter instrument bedoeld voor selectie is de toets steeds meer geworden tot een instrument ter ondersteuning van het onderwijsleerproces. Parallel aan deze ontwikkeling voltrok zich een verandering in de psychometrische analyse van de studietoetsen. Aanvankelijk hadden de toetsen veel meer tot doel in gedetailleerde zin te differentiëren tussen leerlingen. Later, bij gebruik als logisch onderdeel van het onderwijsleerproces, werden de toetsen juist gebruikt om heel globaal te differentiëren tussen leerlingen die wel en leerlingen die niet (volledig) de leerstof beheersten. De psychometrische eisen die men aan de toetsen moet stellen, rekening houdend met het verschillend gebruik ervan, variëren. Voor de toetsen die gebruikt worden als integraal onderdeel van het onderwijsleerproces is m.n. van belang welke grens aangehouden moet worden voor de beslissing voldoende/onvoldoende. Ook de consistentie van deze beslissingen over toetsen is een belangrijk psychometrisch aspect.

In dit hoofdstuk zal m.n. aandacht geschonken worden aan de psychometrische kenmerken van de door ons gebruikte leerstofgebonden toetsen. Het hoofdstuk valt globaal genomen in twee delen uiteen. In het eerste gedeelte wordt ruime aandacht geschonken aan moderne inzichten en procedures aangaande de bepaling van aftestgrensen en de consistentiebepaling van leerstofgebonden toetsen. Het tweede deel van dit hoofdstuk tenslotte is gewijd aan de concrete toepassing van deze technieken op de door ons gebruikte formatieve toetsen uit het eerste en tweede experiment. Toepassing bij de zgn. formatieve hertoetsen blijft achterwege gezien het relatief gering aantal hierbij betrokken proefpersonen.

6.2. Norm-georiënteerd en criterium-georiënteerd meten.

In 1963 schreef de Amerikaan Glaser een inmiddels klassiek artikel (Glaser, 1963), waarin hij twee mogelijke gebruiks-

wijzen van studietoetsen met elkaar vergeleek, nl. norm-georiënteerde en criterium-georiënteerde interpretatie van toetsscores. Afhankelijk van de interpretatie van de scores spreekt men van norm- en criterium-georiënteerde toetsen.

norm-georiënteerde toetsen zijn vooral ontworpen om informatie te verkrijgen over de positie van de leerling m.b.t. zijn leerprestaties ten opzichte van medeleerlingen. De uitspraak die op basis van de toetsresultaten gedaan kan worden betreft een vergelijking, zoals X% van de leerlingen van zijn leeftijd (klas, groep, etc.) is slechter (of beter). Een externe referentiegroep wordt gevormd door de populatie waarmee de leerling op een of meer dimensies overeenkomt.

Norm-georiënteerde toetsen zijn m.n. bedoeld om te kunnen discrimineren tussen goede en slechte leerlingen. De gebruikte scoringsmethoden reflecteren deze intenties. Scores worden veelal gepresenteerd in de vorm van percentielscores, standaardcores, stanines e.d. De toetsconstructie is zodanig dat de relatieve positie van de kandidaten zo betrouwbaar mogelijk kan worden vastgesteld. Teneinde in dit verband maximaal discriminerende toetsen te ontwikkelen, kiest men de items op basis van discriminatie-indices: die items worden opgenomen die goed beantwoord worden door leerlingen die hoog scoren op de gemeten dimensie (tekstbegrip, rekenvaardigheid, o.i.d.) en die fout beantwoord worden door leerlingen die laag scoren op de gemeten dimensie.

Toetsen bedoeld voor selectie zijn in het algemeen de meest zuivere vorm van norm-georiënteerde toetsen.

Bij *criterium-georiënteerde toetsen* gaat het - in tegenstelling tot norm-georiënteerde toetsen - niet om de vergelijking van individuele toetsprestaties met de prestaties van een normgroep. Er kan sprake zijn van een expliciet omschreven en interpreteerbaar continuüm, dat zich uitstrekt van geen kennis of vaardigheid tot perfecte kennis of vaardigheid. De toetsprestatie van een leerling vormt dan een indicatie van de plaats van de leerling op dit continuüm. Bij dgl. toetsen, waarbij het eindgedrag van de leerling indicatief is voor zowel het individuele beheersingsniveau als voor het succes waarmee een volgend leerstofonderdeel kan worden gevolgd, is in feite sprake van een absolute kwaliteitsindex van het eindgedrag. Deze toetsen vormen een integraal deel van het onderwijs, en de interpretatie van de toetsscores vindt plaats op basis van vooraf opgestelde criteria (Glaser, 1973).

Millman (1974) spreekt in dit verband over 'criterion-referenced differential assessment devices'. Conventionele toetsontwikkelingsprocedures zouden nog wel gelden, maar de te toetsen vaardigheid moet voor elke leerling afzonderlijk

interpreteerbaar zijn in termen van taken en vaardigheden zonder directe referentie aan een populatie. Een en ander heeft geleid tot het uitsplitsen van de doelstellingen waarop een onderwijsleerproces is gericht en het gericht willen toetsen in hoeverre de aldus gespecificeerde doelstellingen zijn bereikt door de leerlingen. De doelstellingen omvatten in dit geval een (zeer) beperkt leerstofdomein. De toetsen worden dan ook wel aangeduid als domein-georiënteerd.

Het itemdomein waaruit de toets, middels gestratificeerde random steekproeftrekking, wordt samengesteld, verwijst dan ook direct naar de expliciete onderwijsdoelen. Veelal vindt interpretatie van de toetsresultaten plaats in termen van een dichotome beslissing: de leerling beheerst de leerstof wel, of de leerling beheerst de leerstof niet. In het algemeen is derhalve sprake van een enkelvoudig criterium: scoort de leerling lager dan het criterium dan wordt hij geacht de leerstof niet (voldoende) te beheersen, scoort hij gelijk aan of hoger dan het criterium dan wordt hij geacht de leerstof ook daadwerkelijk te beheersen.

Bij de vaststelling van een criterium kan men op verschillende wijzen te werk gaan, o.m. afhankelijk van het gehanteerde toets-of meetmodel. Meskauskas (1976) onderscheidt continuüm-modellen en 'state-models'.

In continuümmodellen wordt de leerstofbeheersing beschouwd als een continue variabele. Op het beheersingscontinuüm wordt een benedengrens gedefinieerd. Uitgaande van deze benedengrens en het gehanteerde meetmodel destilleert men nu een grensscore op de toets. Als de toetsscore deze grens overschrijdt wordt het beheersingsniveau van de leerling als voldoende beoordeeld. Bij de zgn. 'state-models' gaat men meestal uit van een dichotoom beheersingsniveau. De theoretische leercurve vormt een stapfunctie. De criteriumscore wordt in theorie op 100% gesteld, maar is in werkelijkheid lager vanwege mogelijke veronderstelde meetfouten. Immers, ook leerlingen die de leerstof volledig beheersen zullen fouten maken. Deze fouten worden echter veroorzaakt door storingen binnen de meet-procedure. Vanuit testmethodologisch gezichtspunt concentreert men zich binnen de 'state-models' op het schatten van een domeinscore op grond waarvan de dichotome beslissing onvoldoende/voldoende zodanig kan worden genomen dat het aantal of de gevolgen van foutieve beslissingen tot een (kwantificeerbaar) minimum beperkt worden.

6.3. Criterium-georiënteerde toets- en itemanalyse

Popham en Husek (1969) betogen dat klassieke modellen en

procedures voor toets- en itemanalyse bij criterium-georiënteerde toetsen niet voldoen, omdat de klassieke testtheorie gebaseerd is op verschillen in toetsscores tussen leerlingen, m.a.w. op variantie. Bij criterium-georiënteerde toetsen speelt de scorevariantie echter een ondergeschikte rol, maar gaat het veeleer om de relatie van de scores met een bepaald criterium (Van der Linden, 1980, p.6). De geringe variantie in scores bij het gebruik van criterium-georiënteerde toetsen heeft aanleiding gegeven tot vele voorstellen voor nieuwe betrouwbaarheidscoëfficiënten. Immers de betrouwbaarheidscoëfficiënten binnen de klassieke testtheorie zijn gebaseerd op scorevariantie en dus vaak laag bij criterium-georiënteerde toetsen. Van der Linden (1980, p.7) betoogt echter dat de geringe variantie in toetsscores bij criterium-georiënteerd meten niet impliceert dat de klassieke testtheorie voor deze toetsen tekort schiet. Immers, de klassieke testtheorie bevat, behoudens de eis van eindige variantie, geen assumpties aangaande de grootte van de scorevariantie.

Binnen de zgn. latente trekmodellen (Birnbaum, 1968) heeft men niet alleen te maken met toetsparameters die variantie-onafhankelijk zijn, maar ook, bijv. bij het Rasch-model, onafhankelijk zijn van de populatie waarop de toets wordt afgenomen. Deze zgn. niet-klassieke testtheorie biedt mogelijkheden voor de toets- en itemanalyse van criterium-georiënteerde toetsen.

Een andere aanpak van criterium-georiënteerd meten wordt gevormd door de decisie-theoretische benadering. Hierbij gaat het niet zozeer om de betrouwbaarheid en validiteit van het meetinstrument in klassieke zin, maar om de betrouwbaarheid en validiteit van de *beslissingen* die genomen worden op basis van de toetsscores. Vanuit onderwijskundig gezichtspunt betekent dit dat de toets niet langer beschouwd wordt als bedoeld voor de meting van leerlingprestaties teneinde cijfers of iets dergelijks toe te kennen. Veeleer vervult de toets een functie m.b.t. het nemen van beslissingen aangaande het onderwijsleerproces. De toets vormt, zoals reeds eerder is opgemerkt, een integraal onderdeel van dit proces. In volgende paragrafen van dit hoofdstuk zal bij de evaluatie van de door ons gebruikte formatieve toetsen ook uitgegaan worden van een decisie-theoretische benadering.

6.4. Cesuurbepaling

Het bepalen van de cesuur bij een studietoets is in feite een vorm van classificeren. Er wordt een aantal categorieën

onderscheiden waarin men individuen wil onderbrengen. De veronderstelling is dat deze categorieën latent zijn: het is niet zonder meer aan een leerling te zien in welke categorie hij thuishoort. Teneinde in deze een beslissing te kunnen nemen maken we gebruik van een meetinstrument, waarbij de metingen indicatoren zijn van de latente categorieën. Het is nu de bedoeling deze metingen zodanig te categoriseren dat er een classificatie van individuen ontstaat die op de een of andere wijze optimaal is. Meestal beperkt men het classificatie-probleem tot twee latente categorieën: geschikten/ongeschikten; voldoende/onvoldoende, etc.. Meskauskas (1976) beschrijft een methode om reeds bij de eerste keer de cesuur, of het criterium empirisch te bepalen. Een groep leerlingen die (nog) geen onderwijs m.b.t. een bepaald leerstofonderdeel gekregen hebben moeten een toets maken. Een tweede groep leerlingen maakt dezelfde toets nadat ze m.b.t. de getoetste leerstof onderwijs gevolgd hebben. Het optimale criterium wordt nu gevormd door het snijpunt van de scoreverdelingen der beide groepen. Indien de natoets identiek is met de voortoets kan ook gebruik gemaakt worden van de verdelingen der scores op deze toetsen.

De voorgestelde procedure heeft nogal wat nadelen. De eerste groep, de zgn. nonmasters, is niet duidelijk gedefinieerd. Immers leerlingen die nog nimmer kennis genomen hebben van de betreffende leerstof zullen een geheel andere scoreverdeling (kunnen) hebben dan leerlingen die reeds enige kennis van het onderwerp hebben, of reeds enigermate kennis hebben gemaakt met de leerstof. Voorts worden de zgn. masters in het algemeen niet gedefinieerd als leerlingen die beter scoren dan leerlingen die de leerstof (nog) helemaal niet kennen. Bovendien is de validering van het criterium niet gelegen in het onderwijs dat aan de toets voorafgaat, maar juist in het succes op onderwijs dat volgt op en dat weer wordt bepaald door de toetsscore van die les.

Een andere procedure waarbij slechts één toetsafname noodzakelijk is, is het schatten van domeinscores op basis van geobserveerde scores. Millman (1974) beschrijft een binomiaal en een Bayesiaans model voor de schattingsprocedure. In beide modellen worden dichotome scoring en locale onafhankelijkheid van de antwoorden verondersteld. Bij de schatting van de domeinscore via het binomiaal model wordt alleen rekening gehouden met de prestatie van het individu op de toets. Met behulp van tabellen kan vervolgens de kans op een juiste beslissing berekend worden, gegeven een criteriumscore en een geschatte domeinscore. In het Bayesiaanse model wordt additionele informatie gebruikt die de leerkracht heeft over bijv. de groep waartoe de leerling behoort, of over eerdere

prestaties van de leerling, etc.. Het resultaat is de kans welke een domeinscore heeft om boven een bepaalde toetsscore te liggen.

6.4.1. Utiliteiten

In de voorgaande paragraaf is in algemene zin aangegeven wat onder cesuurbepaling verstaan dient te worden. Meer geformaliseerd hebben we te maken met het volgende (cf. Veldhuijzen, z.j., p.15):

- er is een unidimensionele dispositie waar men scores θ op behaalt;
- er is een criteriumscore θ_0 , die de categorieën specificiert $\theta < \theta_0$ en $\theta > \theta_0$;
- er is een toetsscore x , die via een bepaald meetmodel indicatief is voor θ ;
- er is een positieve samenhang tussen x en θ ;
- men zoekt een grensscore x_0 met het oogmerk individuen met $x > x_0$ te classificeren als: $\theta > \theta_0$ en individuen met $x < x_0$ als $\theta < \theta_0$;
- men wenst x_0 zo te kiezen dat de classificatie optimaal is.

Het probleem van de cesuurbepaling heeft nu alles te maken met wat onder 'optimale classificatie' verstaan moet worden. Immers het classificeren van leerlingen als geslaagd of gezakt zal niet zonder fouten geschieden, daar het gebruikte meetmodel over het algemeen stochastisch van aard is.

Op grond van de toetsscore $x > x_0$ van een leerling beslist men dat de leerling geslaagd is, dus $\theta > \theta_0$. In werkelijkheid kan de geobserveerde score x door toevallige factoren groter dan x_0 zijn, terwijl in werkelijkheid $\theta < \theta_0$ is. Met andere woorden: de leerling is ten onrechte geslaagd. Hoe ernstig men een dergelijke misclassificatie vindt, wordt bepaald door de beslisser bij het indiceren van θ_0 en x_0 .

In algemene zin kan men nu spreken van het nut, de opbrengst of de *utiliteit* van een uitkomst. Utiliteiten zijn schaalwaarden voor al of niet subjectieve aspecten, zoals de subjectieve beleving van de leerling om een les te moeten herhalen, of - na het zakken voor een examen - een heel studiejaar te moeten overdoen.

In de bovenbeschreven situatie kunnen we vier uitkomsten onderscheiden:

1. gegeven: $\theta < \theta_0$; waargenomen $x < x_0$; beslissing: de leerling is ongeschikt (juiste beslissing);
2. gegeven: $\theta < \theta_0$; waargenomen $x > x_0$; beslissing: de

- leerling is geschikt (foute beslissing);
3. gegeven: $\theta > \theta_0$; waargenomen $x \geq x_0$; beslissing: de leerling is geschikt (juiste beslissing);
4. gegeven: $\theta > \theta_0$; waargenomen $x < x_0$; beslissing: de leerling is ongeschikt (foute beslissing).

Elk van de vier bovenstaande resultaten heeft in een leerlingenpopulatie een kans om voor te komen. De beslisser (bijv. de leerkracht) verbindt aan elk der vier mogelijke uitkomsten een bepaalde waarde, of utiliteit. Deze utiliteiten, kansen en beslissingen kunnen samengevat worden in een matrix (fig. 6.1):

lat. variabele	Observ. ----- besliss.	$x < x_0$	$x \geq x_0$
		$\theta < \theta_0$	$\theta \geq \theta_0$
$\theta \geq \theta_0$		p_{12} u_{12}	p_{11} u_{11}
$\theta < \theta_0$		p_{22} u_{22}	p_{21} u_{21}

Fig. 6.1. Matrix van kansen en utiliteiten

In figuur 6.1. is de som der kansen p_{ij} vanzelfsprekend gelijk aan 1.

In het kader van de cesuurbepaling gaat het er nu om x_0 zodanig te kiezen dat steeds de beslissing met de grootste utiliteit genomen wordt. In de praktijk is dit vanzelfsprekend niet zonder meer mogelijk, daar de latente variabele θ niet waarneembaar is. Voor een gegeven cesuur x_0 kunnen nu in een leerlingenpopulatie de volgende grootheden bepaald worden:

- $\phi_1(x_0)$: de verwachte utiliteit van alle te nemen beslissingen, gegeven dat $\theta \geq \theta_0$;
- $\phi_2(x_0)$: de verwachte utiliteit van alle te nemen beslissingen, gegeven dat $\theta < \theta_0$.

Vanuit decisie-theoretisch gezichtspunt kan men nu via twee methoden het beslissingsproces optimaliseren,

1. Kies x_0 , dat de verwachte utiliteit van alle te nemen beslissingen - dus de grootheid $\sum_{i=1}^2 \sum_{j=1}^2 u_{ij} p_{ij}$ - zo groot mogelijk is:

maximaliseer de verwachte waarde van $\phi_1(x_0)$ en $\phi_2(x_0)$.

2. Kies x_0 zodanig dat de utiliteit van de beslissing met de kleinste utiliteit van allemaal, nog zo groot mogelijk is (m.a.w. maximaliseer het minimum van $\phi_1(x_0)$ en $\phi_2(x_0)$).

In formele zin kunnen beide procedures als volgt genoteerd worden:

$$(1) \text{ 'Bayes-utiliteit' : } \max_{x_0} \{ \phi_1(x_0)P(\theta > \theta_0) + \phi_2(x_0)P(\theta < \theta_0) \}$$

$$(2) \text{ 'Maximin-utiliteit': } \max_{x_0} \{ \min [\phi_1(x_0), \phi_2(x_0)] \}$$

De eerste methode is de in de psychometrie welbekende 'Bayes-procedure', de tweede methode is de zgn. 'maximin-procedure'. Laatstgenoemde is in het kader van een decisie-theoretische benadering beschreven door Veldhuijzen (z.j.; 1979, 1982). Ook Huynh (1980) geeft in Psychometrika een uitvoerige beschrijving.

6.5. Betrouwbaarheidsbepaling

Hambleton, Swaminathan, Algina & Coulson (1978) onderscheiden m.b.t. criterium-georiënteerd meten drie verschillende soorten betrouwbaarheid:

1. betrouwbaarheid van classificatiebeslissingen m.b.t. de leerstofbeheersing;
2. betrouwbaarheid van criterium-georiënteerde toetsscores;
3. betrouwbaarheid van de schattingen der domein-score.

In het kader van deze studie is vooral de eerste vorm van betrouwbaarheid, namelijk die der classificatiebeslissingen, van belang. Beter is het eigenlijk om in dit geval niet te spreken van betrouwbaarheid, maar van consistentie der classificatiebeslissingen (gezakt/geslaagd) indien dezelfde groep herhaalde malen de toets afgenomen zou worden (cf. Van der Linden, 1981).

Subkoviak (1980) geeft een overzicht van vijf in de literatuur te vinden methoden voor de consistentiebepaling van classificatiebeslissingen aangaande het leerstofbeheersingsniveau van leerlingen. Elk dezer methoden zal in het navolgende kort besproken en geëvalueerd worden.

* Opgemerkt moet worden dat men om θ_1 en θ_2 te bepalen niet altijd een groep leerlingen nodig heeft. Met één leerling kan het ook. In het beta-binomiale model echter - en dit model gebruiken we hier - is wel sprake van een verdeling van θ over leerlingen.

6.5.1. De methode van Carver

In de door Carver (1970) voorgestelde methode worden twee paralleltoetsen bij dezelfde groep afgenomen, of dezelfde toets bij twee vergelijkbare groepen. Vervolgens worden de percentages leerlingen die op beide toetsafnames als geslaagd beschouwd worden met elkaar vergeleken. Als de percentages (vrijwel) gelijk zijn worden de toetsen als betrouwbaar gekwalificeerd.

Een nadeel van deze methode is dat ook - in termen van de klassieke testtheorie - onbetrouwbare toetsen vergelijkbare percentages kunnen opleveren; terwijl bovendien de consistentie (of gebrek daaraan) van individuele classificaties buiten beschouwing blijft.

6.5.2. De methode van Swaminathan-Hambleton & Algina

In deze methode wordt eveneens gebruik gemaakt van twee toetsafnames. Nu echter wordt wel rekening gehouden met individuele misclassificaties. Bovendien vindt een correctie plaats voor de proportie consistente beslissingen die op basis van toeval te verwachten zijn. Als index voor de consistentie van de beslissingen wordt de kappa-coëfficiënt van Cohen (1960) aanbevolen.

6.5.3. De methode van Huynh

Huynh (1976) gaat uit van de praktijksituatie waarin het in het algemeen onmogelijk is herhaalde malen dezelfde toets af te nemen. Hij ontwikkelt dan ook een procedure die de betrouwbaarheid van de classificatiebeslissing schat op basis van één toetsafname (Huynh spreekt over betrouwbaarheid in plaats van consistentie). Uitgangspunt of aanname bij de onderstaande procedure is, dat de items volgens de een of andere aselechte steekproeftrekkingsprocedure afkomstig zijn uit een welomschreven (leerstof/items)-domein.

6.5.3.1. Het bivariate beta-binomiale model

Voor toetsen die betrekking hebben op een relatief beperkt kennis- of vaardigheidsdomein blijkt het beta-binomiale model zeer bruikbaar te zijn (Lord & Novick, 1968). Bij een toets bestaande uit n items, een toetsscore x voor een leerling met relatieve ware score θ (de proportie items in het itemuniversum

of -domein, die de leerling kent/beheerst) geldt dat de toetsscore x een binomiaalverdeling heeft, namelijk

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \dots, n \text{ en } 0 \leq \theta \leq 1.$$

Het binomiaal model gaat er van uit dat alle items in het universum uitwisselbaar ('exchangeable') zijn. Met andere woorden items zijn uitwisselbaar als de simultane scoreverdeling invariant is onder permutaties van de items. Huynh (1976, p. 254) merkt op dat voor toetsen die een integraal onderdeel vormen van het onderwijsleerproces (zoals de door ons gebruikte formatieve toetsen), de binomiale verdeling een alleszins redelijke benadering vormt van de toetsscoreverdeling. De ware score θ volgt een beta-verdeling (cf. Mood, Graybill & Boes, 1974, p. 115):

$$f(\theta; \alpha, \beta) = B^{-1}(\alpha, \beta) \theta^{\alpha-1} (1-\theta)^{\beta-1} \text{ voor: } 0 < \theta < 1. \\ = 0 \text{ voor alle andere waarden van } \theta.$$

De parameters α en β zijn resp. (cf. Lord & Novick, 1968, p. 517-520):

$$\alpha = (-1 + 1/\alpha_{21})\mu$$

$$\text{en } \beta = -\alpha + n/\alpha_{21} - n,$$

$$\text{waarbij } \alpha_{21} = \frac{n}{n-1} \left[1 - \frac{\mu(n-\mu)}{n\sigma^2} \right]$$

en μ het gemiddelde en σ de standaardafwijking zijn van de verdeling van toetsscores in de populatie.

De marginale verdeling van de toetsscore x is de zgn. negatief-hypergeometrische, of beta-binomiale verdeling van de vorm

$$f(x) = \binom{n}{x} B(\alpha+x, n+\beta-x) / B(\alpha, \beta),$$

waarbij B de beta-functie aanduidt, die gedefinieerd is als:

$$B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta.$$

Voor waarden van x in de buurt van 0, alsmede voor waarden in de buurt van n geeft Huynh (1976, p. 254) twee berekeningsformules. Indien we in plaats van over één de beschikking hebben over twee equivalente criterium-georiënteerde toetsen kunnen we de simultane verdeling van de toetsscores x en y bepalen. We gaan ervan uit dat we, althans in theorie, een oneindig aantal equivalente toetsen zouden kunnen samenstellen o.g.v. aselechte steekproeven uit het itemdomein. Bovendien

zijn de waargenomen toetsscores onafhankelijk van de toevallig gebruikte toets: behoudens meetfouten zouden we bij gebruik van een andere equivalente toets dezelfde toetsscore waarnemen. Indien aan beide condities voldaan wordt, dan zijn de scores verkregen op equivalente toetsen lokaal onafhankelijk. Indien nu x en y de scores zijn op de equivalente toetsen X en Y , dan geldt onder aanname van lokale onafhankelijkheid dat x en y een bivariate beta-binomiale (of negatief hypergeometrische) verdeling volgen met als verdelingsfunctie:

$$f(x,y) = \frac{\binom{n}{x} \binom{n}{y}}{B(\alpha, \beta)} B(\alpha+x+y, 2n+\beta-x-y).$$

Op analoge wijze als bij de univariate beta-binomiale verdeling geeft Huynh ook hiervoor twee berekeningsformules voor de waarden van x en y in de buurt van 0 resp. n .

6.5.3.2. Berekening en evaluatie van κ (coëfficiënt kappa)

Bij afname van twee equivalente toetsen kan onderstaande beslissingstabel geconstrueerd worden, waarbij geldt dat de marginalen gelijk zijn (dus $p_{0.} = p_{.0}$ en $p_{1.} = p_{.1}$).

Toets X \ Toets Y	gezakt	geslaagd	marginale proportie
gezakt	p_{00}	p_{01}	$p_{0.}$
geslaagd	p_{10}	p_{11}	$p_{1.}$
marginale proportie	$p_{.0}$	$p_{.1}$	$p_{..}$

Fig. 6.2. Classificatietabel o.g.v. 2 equivalente toetsen

De proportie leerlingen die op beide toetsen een score gelijk aan of hoger dan de criteriumscore (cesuur) behalen, wordt weergegeven door p_{11} . De proportie leerlingen die zo'n score slechts op een der twee toetsen behalen, is gelijk aan $p_{01} + p_{10}$. Er geldt nu dat:

$$p_{11} = \sum_{x=c}^n \sum_{y=c}^n f(x,y)$$

en

$$p_{1.} = \sum_{x=c}^n f(x).$$

De verwachte proporties kunnen berekend worden m.b.v. de in de vorige paragraaf genoemde berekeningsformules voor de uni- en bivariate beta-biominale verdeling.

Aangetoond kan worden* dat door onderstaande formule de betrouwbaarheidsindex kappa gedefinieerd wordt:

$$\kappa = (p_{11} - p_{1.}^2) / (p_{1.} - p_{1.}^2).$$

De numerieke waarden van kappa liggen tussen 0 en 1. (Huynh (1978) toont aan dat, als x_1 en x_1' uitwisselbaar zijn, $p_{11} > p_{1.}^2$, waardoor $\kappa > 0$).

Tot nu toe is telkens gesproken over twee equivalente toetsen. Het is echter ook mogelijk een schatting te maken van kappa o.g.v. één toetsafname.

Hiertoe worden voor de berekening van de parameters α en β van de beta-verdeling (zie blz. 82) de grootheden μ en σ vervangen door de steekproefwaarden \bar{x} en s , waarbij n het aantal toets-items is. Vervolgens wordt $p_1 = f(n) + f(n-1) + \dots + f(c)$ berekend m.b.v. de formules (Lord & Novick, 1968, p. 515)

*De overeenstemmingscoëfficiënt κ wordt gedefinieerd (cf. Cohen, 1960) als:

$$\kappa = \frac{p - p_c}{1 - p_c}, \text{ waarbij } p = \sum_{k=1}^m p_{kk}$$

$$p_c = \sum_{k=1}^m p_{k.} \cdot p_{.k}$$

Er geldt nu (fig. 6.2.):

$$p = p_{00} + p_{11} = (p_{0.} - p_{01}) + p_{11} = \{(1 - p_{1.}) - (p_{.1} - p_{11})\} + p_{11} =$$

$$= 1 - 2(p_{.1} - p_{11})$$

$$p_c = p_{0.}^2 + p_{1.}^2 = (1 - p_{1.})^2 + p_{1.}^2 = 1 - 2(p_{1.} - p_{1.}^2)$$

$$p - p_c = (1 - 2p_{.1} + 2p_{11}) - (1 - 2p_{1.} + 2p_{1.}^2) = 2(p_{11} - p_{1.}^2)$$

$$1 - p_c = 1 - \{1 - 2(p_{1.} - p_{1.}^2)\} = 2(p_{1.} - p_{1.}^2)$$

$$\kappa = (p_{11} - p_{1.}^2) / (p_{1.} - p_{1.}^2)$$

$$f(n) = \prod_{i=1}^n \frac{n+\alpha-1}{n+\alpha+\beta-i}$$

en

$$f(x-1)=f(x) \cdot \frac{x(n+\beta-x)}{(n-x+1)(\alpha+x-1)}, \text{ waarbij } x=1, \dots, n.$$

De proportie $p_{11} = f(n,n)+f(n,n-1)+f(n-1,n)+\dots+f(c,c)$ wordt berekend middels de formules:

$$f(n,n) = \prod_{i=1}^{2n} \frac{2n+\alpha-i}{2n+\alpha+\beta-i}$$

en

$$f(x-1,y)=f(x,y) \cdot \frac{x(2n+\beta-x-y)}{(n-x+1)(\alpha+x+y-1)}.$$

Tenslotte worden de beide berekende proporties ingevuld in de bovenaangeduide berekeningsformule voor de kappa-coëfficiënt. In de voorafgaande procedure werden o.g.v. één toetsafname de parameters α en β geschat, waarbij zeer acceptabele schattingen verkregen kunnen worden met relatief weinig proefpersonen (N): bij zeer korte toetsen $N < 40$ of $N < 2n$ bij langere toetsen (cf. Subkoviak, 1980, p.139). De nogal gecompliceerde berekeningen kunnen middels een computerprogramma verricht worden. Ook is er een benaderingsprocedure.

6.5.4. De methode van Subkoviak

Subkoviak (1976) definieert de betrouwbaarheidscoëfficiënt voor leerling i ('coefficient of agreement') als de waarschijnlijkheid dat i op grond van zijn scores op de paralleltoetsen X en X' in dezelfde beheersingsklasse (geslaagd/gezakt) geplaatst wordt. De betrouwbaarheidscoëfficiënt $P_C(i)$ voor leerling i bij criterium C kan nu geschreven worden als:

$$P_C(i) = P(X_i \geq C, X'_i \geq C) + P(X_i < C, X'_i < C).$$

Het rechterlid van bovenstaande vergelijking geeft de waarschijnlijkheid van consistente geslaagd/geslaagd- en gezakt/gezakt-beslissingen. De betrouwbaarheidscoëfficiënt P_C voor een groep van N leerlingen bij criterium C wordt gedefinieerd als het gemiddelde van de coëfficiënten van individuele leerlingen, derhalve:

$$P_C = \frac{1}{N} \sum_{i=1}^N P_C(i).$$

Teneinde de individuele, alsmede de groepsbetrouwbaarheidscoëfficiënt te kunnen schatten is het, evenals bij de methode van Huynh, nodig enige assumpties in te voeren. De eerste assumptie is dat de scores X_i en X'_i onafhankelijk verdeeld zijn voor leerling i (Lord & Novick, 1968). De betrouwbaarheidscoëfficiënt voor één leerling kan dan geschreven worden als:

$$P_c^{(1)} = P(X_i > C) \cdot P(X'_i > C) + P(X_i < C) \cdot P(X'_i < C)$$

(zie Huynh, 1976, p. 267-8).

De tweede assumptie houdt in dat de verdelingen van X_i en X'_i voor een bepaalde leerling een identieke binomiale vorm hebben, waarbij de volgende assumpties gelden:

1. elk der n toetsitems wordt 0 of 1 gescoord;
2. het resultaat op het ene item beïnvloedt het resultaat op een ander item niet;
3. de kans op een goed antwoord blijft constant over de items.

De betrouwbaarheidscoëfficiënt wordt dan:

$$P_c^{(i)} = [P(X_i > C)]^2 + [1 - P(X_i > C)]^2$$

waarin

$$P(X_i > C) = \sum_{X_i=C}^n \binom{n}{X_i} p_i^{X_i} (1-p_i)^{n-X_i}.$$

De grootheid p_i in bovenstaande verdelingsfunctie is de kans dat een item goed beantwoord wordt door leerling i . Bij de schatting van p_i o.g.v. een toets met een relatief gering aantal items kunnen betere schattingen verkregen worden door rekening te houden met collaterale informatie, zoals het gemiddelde en de variantie van de groep leerlingen waarbij de toets is afgenomen. Als de verdeling van de waargenomen scores X in de groep een negatief hypergeometrische verdeling benadert, dan is onderstaande regressievergelijking een goede schatting van p_i (Lord & Novick, 1968, p. 516):

$$\hat{p}_i = \alpha_{21} \frac{X_i}{n} + (1 - \alpha_{21}) \frac{M_x}{n} \quad (\text{waarbij } M_x = \text{het gemiddelde der waargenomen scores})$$

waarin

$$\alpha_{21} = \frac{n}{n-1} \left[1 - \frac{M_x (n - M_x)}{n S_x^2} \right] \quad (\text{waarbij } S_x^2 = \text{de variantie der waargenomen scores})$$

de zgn. KR-21 betrouwbaarheidscoëfficiënt is.

Met behulp van bovenstaande formules is het mogelijk om op basis van één toetsafname een schatting te maken van de (beslissings) betrouwbaarheid. Bij de afleiding van de diverse formules is uitgegaan van het binomiale model. De aanname daarbij is o.m. dat de kans op een goed antwoord constant blijft over de items. Gaan we echter uit van de -realistischer- situatie dat de kans op succes varieert over de items dan kan men i.p.v. het enkelvoudige binomiale model uitgaan van het samengestelde binomiale model, terwijl in de regressievergelijking van p_i gebruik gemaakt moet worden van de KR-20 betrouwbaarheidscoëfficiënt i.p.v. de KR-21.

6.5.5. De methode van Marshall & Haertel

Subkoviak (1976, 1980) beschrijft tenslotte een methode van Marshall & Haertel waarin eveneens op basis van één toetsafname een schatting gemaakt wordt van de beslissings-betrouwbaarheid. De aanname hierbij is dat als een leerling herhaalde malen getoetst wordt, de waargenomen scores binomiaal verdeeld zijn.

Op basis van de scores op een toets bestaande uit n items schatten Marshall & Haertel de scores van de groep op een hypothetische toets bestaande uit $2n$ items. Theoretisch kunnen we nu deze toets opsplitsen in halve toetsen van elk n items en vervolgens kan dan een schatting gemaakt worden van de proportie leerlingen die consistent geclassificeerd worden op grond van de twee halve toetsen.

Een dgl. schatting kan verkregen worden voor elk van de mogelijke opsplitsingen van de toets bestaande uit $2n$ items. Het gemiddelde van al deze 'split-half'-coëfficiënten vormt de uiteindelijke schatting van p_0 , de proportie consistente classificaties op de twee toetsen.

6.5.6. Vergelijking en evaluatie der methodes

De methoden van Carver en Swaminathan-Hambleton-Algina hebben in vergelijking met de drie andere methoden als zeer belangrijk nadeel dat voor de betrouwbaarheidsbepaling (in de zin van consistentie van classificaties) twee toetsafnames vereist zijn.

We zullen deze methoden dan ook hier verder buiten beschouwing laten.

Uit een empirisch onderzoek van Subkoviak (1978) blijkt dat de methoden van Huynh, Subkoviak en Marshall-Haertel consistentie-schatters opleveren met relatief geringe

standaardmeetfouten voor kleine steekproeven van plm. 30 leerlingen. Nadelen van de methodes zijn dat ze rekentechnisch nogal bewerkelijk zijn en voor zeer korte toetsen (< 10 items) onzuivere ('biased') schatters opleveren. De methode van Huynh geeft bij kleine toetsen een onderschatting op alle criteriumniveaus. De methode van Subkoviak daarentegen geeft onderschattingen voor criteriumniveaus rond het centrum van de scoreverdeling (50%) en geringe overschattingen voor criteria in de staarten van de verdeling (bijv. bij 80%). Bij de methode van Marshall-Haertel is dit precies het omgekeerde. Voor de consistentiebepaling van de door ons gebruikte formatieve toetsen zullen we uitgaan van de methodes van Huynh en Subkoviak, aangezien in de literatuur deze procedures ook de meeste aandacht krijgen.

De twee betrouwbaarheidscoëfficiënten κ en p_0 die door deze methodes verkregen worden zijn niet identiek. Zij meten verschillende vormen van consistentie.

Coëfficiënt p_0 heeft betrekking op de totale proportie consistente classificaties die tot stand komen op twee toetsen om welke reden dan ook. Factoren die bijdragen tot een redelijke mate van consistentie zoals gemeten door p_0 zijn bijv. het relatieve aantal geslaagden/gezakten in de groep leerlingen waaraan de toets is afgenomen, alsmede de precisie of nauwkeurigheid van de toets. Coëfficiënt κ daarentegen heeft betrekking op consistentie als gevolg van de toets. Indien we p_0 'corrigeren' voor de consistentie ten gevolge van de specifieke geslaagd/gezakt-samenstelling van de groep dan houden we datgene over wat door coëfficiënt κ gemeten wordt.

6.6. Psychometrische analyse formatieve toetsen experiment I

In het voorafgaande is aandacht besteed aan cesuur- en consistentiebepaling bij criterium-georiënteerde toetsen. In het navolgende zullen de besproken procedures toegepast worden op de in experiment I gebruikte formatieve, of criterium-georiënteerde toetsen.

6.6.1. Modeltoetsing

Zowel bij de cesurbepaling als bij de berekening van consistentie-coëfficiënten wordt uitgegaan van de assumptie dat de toetsscores een negatief-hypergeometrische verdeling volgen. Het verdient derhalve aanbeveling om na te gaan of de

waargenomen toetsscoreverdelingen in redelijke mate voldoen aan deze assumptie. Met behulp van de χ^2 -toets werd getoetst op het significant zijn van de verschillen tussen waargenomen aantallen waarnemingen verdeeld over k klassen, en de theoretisch te verwachten aantallen in die k klassen. Het aantal vrijheidsgraden is dan gelijk aan k-3. In tabel 6.1 zijn de resultaten van de significantie-toetsing vermeld.

Tabel 6.1. Modeltoetsing formatieve toetsen experiment I

Formatieve toets/ df			χ^2	Formatieve toets/ df			χ^2
subtoets				subtoets			
• Toets 1	12		15.39	• Toets 4	12		15.51
subtoets 1.1	4		8.46	subtoets 4.1	6		10.73
subtoets 1.2	3		4.20	subtoets 4.2	5		11.67*
subtoets 1.3	3		.49				
• Toets 2	6		10.47	• Toets 5	-		- ¹⁾
subtoets 2.1/2.2	3		2.97	subtoets 5.1	24		13.15
subtoets 2.3/2.4	5		7.26	subtoets 5.2	3		2.84
				subtoets 5.3	7		5.74
• Toets 3	8		9.35	• Toets 6	9		13.23
subtoets 3.1	4		3.38	subtoets 6.1	2		1.38
subtoets 3.2	3		.83	subtoets 6.2	6		11.05*
				• Toets 7	1		9.52*

* $p < .05$

** $p < .01$

¹⁾ geen berekening i.v.m. beperking computerprogramma t.a.v. het aantal items.

Zoals uit tabel 6.1 blijkt, wijken alleen de geobserveerde verdelingen van subtoets 4.2 en toets 7 significant af van de theoretisch verwachte negatief-hypergeometrische verdeling. Hoe ernstig deze discrepantie is, kan blijken uit fig. 6.3. In ieder geval moet bij deze toetsen voorzichtigheid betracht worden bij de interpretatie van de cesuren en consistentie-coëfficiënten, aangezien waarschijnlijk niet voldaan is aan de assumptie over de vorm van de scoreverdeling. In fig. 6.3 is voor alle toetsen aangegeven in welke mate de geobserveerde toetsscores afwijken van de theoretisch verwachte scores. Vaak geeft men deze afwijkingen grafisch weer door de theoretische verdeling en de geobserveerde verdeling in één figuur te tekenen. Alle afwijkingen van de waargenomen scores t.o.v. de verwachte scores worden op het oog even zwaar gewogen. Men kan

echter stellen dat het minder ernstig is als men minder waarnemingen doet dan volgens het model te verwachten zijn, dan dat men waarnemingen doet terwijl dat volgens het model niet 'mag'. (cf. Hays, 1969, p. 582)

In het eerste geval kan men een en ander nog toeschrijven aan de relatief beperkte omvang van de steekproef: bij een grotere steekproef had men wellicht meer waarnemingen - conform het model - gedaan. In het tweede geval kan men de waarneming gegeven het model eigenlijk niet verklaren.

We beelden per toetsscore de geconstateerde afwijking t.o.v. een lijn af, die aangeeft dat er geen discrepantie bestaat tussen geobserveerde en verwachte score-verdeling. In de grafiek wordt de toetsscore op de horizontale as afgezet en op de verticale as $\chi(x)$, waarbij

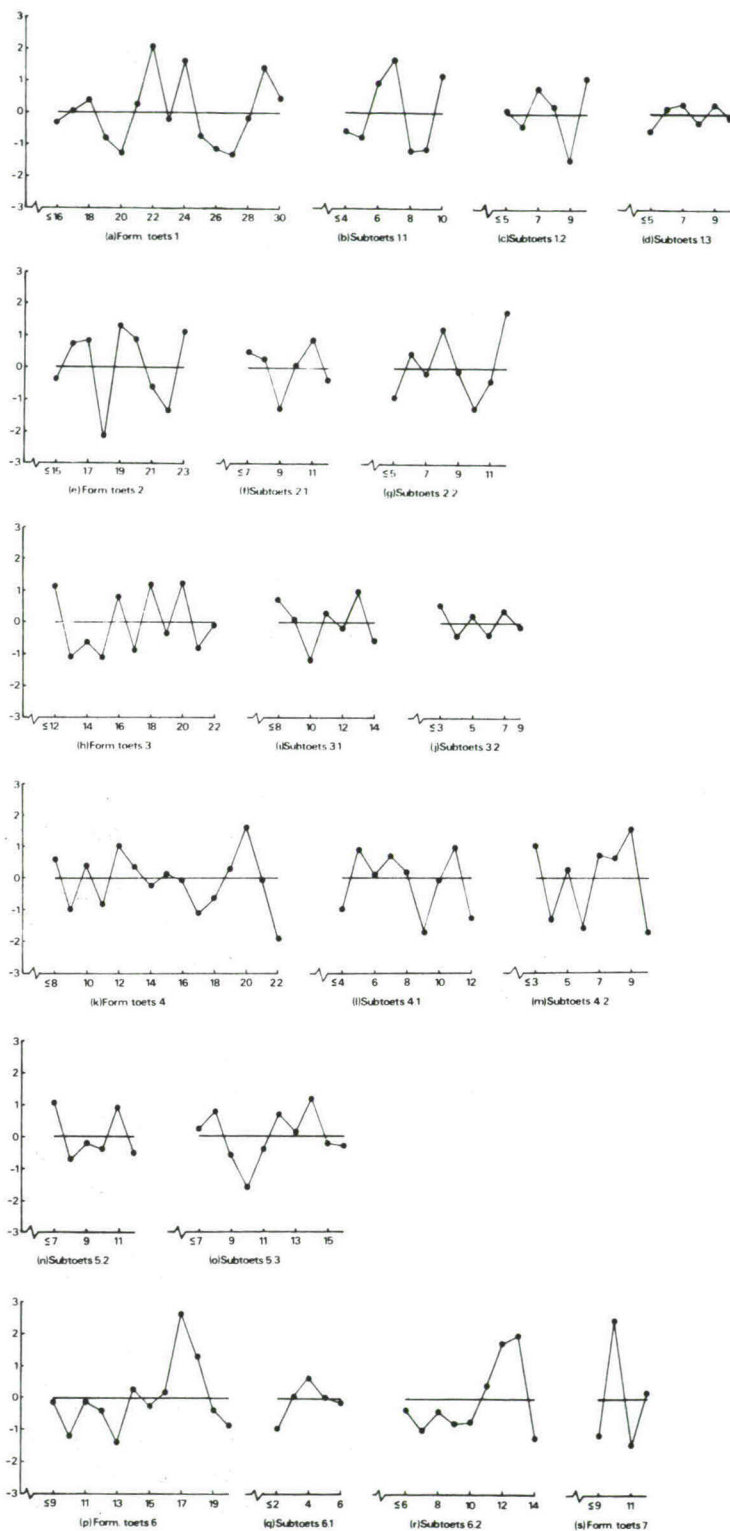
$$\chi(x) = \frac{f_o(x) - f_e(x)}{\sqrt{f_e(x)}} \quad \begin{array}{l} (f_o(x): \text{de geobserveerde frequentie per} \\ \text{toetsscore} \\ f_e(x): \text{de verwachte frequentie per} \\ \text{toetsscore}). \end{array}$$

Indien de waargenomen frequentieverdeling samenvalt met de verwachte frequentieverdeling dan zijn er geen afwijkingen en zal de grafiek slechts de horizontale nullijn vertonen.

In fig. 6.3. (a t/m s) zijn voor alle toetsen de geconstateerde afwijkingen t.o.v. de nullijn ('geen afwijking') afgebeeld. In de formule voor $\chi(x)$ kan geconstateerd worden dat waarnemingen, gedaan waar ze volgens het model niet verwacht worden, ertoe leiden dat de absolute waarde van $\chi(x)$ sterk toeneemt. Hiermee wordt de ernst van deze afwijking dus ook numeriek en grafisch zichtbaar gemaakt.

Uit de grafieken voor subtoets 4.2 (fig. 6.3. (m)) en toets 7 (fig. 6.3. (s)) kunnen we opmaken dat afwijkingen t.o.v. de theoretische verdeling m.n. optreden in de lagere scores: daar waar het model geen waarnemingen verwacht vinden wij ze toch. Het patroon dat uit de diverse grafieken naar voren komt, geeft geen aanleiding te veronderstellen dat een ander dan een aselekt karakter ten grondslag ligt aan de waargenomen afwijkingen.

Het lijkt derhalve niet onredelijk om bij de bepaling van de aftestgrenzen en de consistentie-coëfficiënten ervan uit te gaan dat aan de assumptie m.b.t. de negatief-hypergeometrische verdeling voor de formatieve toetsen en subtoetsen voldaan is. Een uitzondering moet gemaakt worden voor subtoets 4.2 en



Figuur 6.3. Geconstateerde afwijkingen van de toetsscores (exp. I)

toets 7. Voor toets 5 en subtoets 5.1 konden geen grafieken getekend worden i.v.m. de voor deze toetsen te beperkte omvang van het computerprogramma. Overigens moet wel bedacht worden, dat in totaal 19 χ^2 -toetsingen zijn uitgevoerd, waarvan er twee significant bleken. Zelfs bij onafhankelijke toetsingen is iets dergelijks onder aanname van de nulhypothese wel mogelijk. In dit geval zijn voorts niet alle toetsingen onafhankelijk, zodat de resultaten alleszins redelijk te noemen zijn.

6.6.6. Cesuurbepaling formatieve toetsen experiment I

In tabel 6.2 zijn enige beschrijvende gegevens vermeld betreffende de in het eerste experiment gebruikte formatieve toetsen en subtoetsen.

In de tweede kolom is het aantal proefpersonen vermeld dat de toetsen heeft gemaakt. Geen rekening is gehouden met de diverse in het experiment gehanteerde condities. Gezien het feit dat lessen en toetsen voor alle proefpersonen identiek waren lijkt daar ook geen aanleiding toe te zijn.

De derde kolom bevat het aantal items per toets/subtoets, de volgende kolommen vermelden resp. het gemiddelde (M), de standaardafwijking (SD), de homogeniteits- of betrouwbaarheids-coëfficiënt KR-20, alsmede de standaardmeetfout (SE).

Geconstateerd kan worden dat het gemiddelde op vrijwel alle toetsen en subtoetsen ongeveer overeenstemt met de gehanteerde 80%-cesuur (zie onder), terwijl ook de standaardmeetfout binnen alleszins redelijke grenzen blijft (behalve daar waar de KR-20 relatief laag is). De relatief hoge (gemiddelde) scores behoeven geen bevreemding te wekken gezien de direct bij de leerstof(doelstellingen) aansluitende toetsen, en het aan de toetsen voorafgaande onderwijsleerproces. Bij de in onze experimenten gebruikte criterium-georiënteerde toetsen moest reeds vóórdat de proefpersonen een toets gemaakt hadden een beslissing genomen worden aangaande de grens onvoldoende/voldoende. In het onderzoek hebben we deze grens gesteld op 80% van het te behalen aantal punten op een toets. Dat wil zeggen bij een toets van 20 items diende de proefpersoon minimaal 16 items goed te beantwoorden om zijn prestatie als 'voldoende' gekwalificeerd te krijgen. De keuze van de 80%-grens, werd gebaseerd op empirische evidentie zoals gerapporteerd door Block (1972). Onderzoek had o.m. aangetoond dat wanneer men de cesuur op 70% stelde, teveel leerlingen later in de leerstof vastliepen: kennelijk beheersten ze de voorafgaande leerstof toch onvoldoende. Stelde men daarentegen

Tabel 6.2. Descriptieve gegevens formatieve toetsen
experiment I

Formatieve toets/ subtoets	N	k	M	SD	KR-20	SE
Toets 1	84	30	24.64	3.19	.65	1.89
subtoets 1.1	84	10	8.06	1.49	.43	1.13
subtoets 1.2	84	10	7.94	1.34	.27	1.14
subtoets 1.3	84	10	8.64	1.25	.35	1.01
Toets 2	84	24	19.52	2.00	.31	1.66
subtoets 2.1/2.2	84	12	10.77	1.33	.45	.99
subtoets 2.3/2.4	84	12	8.75	1.60	.34	1.30
Toets 3	84	22	19.23	2.48	.68	1.40
subtoets 3.1	84	14	12.29	1.55	.50	1.10
subtoets 3.2	84	8	6.94	1.22	.48	.88
Toets 4	84	22	16.86	3.30	.74	1.68
subtoets 4.1	84	12	9.32	1.83	.56	1.21
subtoets 4.2	84	10	7.54	1.84	.60	1.16
Toets 5	84	78	64.77	7.52	.85	2.92
subtoets 5.1	84	50	40.39	6.00	.84	2.41
subtoets 5.2	84	12	10.25	1.28	.33	1.05
subtoets 5.3	84	16	14.13	2.30	.75	1.14
Toets 6	84	20	18.00	2.79	.83	1.17
subtoets 6.1	84	6	5.18	1.08	.52	.75
subtoets 6.2	84	14	12.82	1.89	.77	.91
Toets 7	84	12	11.71	.63	.33	.51

de cesuur op 90% dan werkte dit voor teveel leerlingen demotiverend, waardoor eveneens te weinig leerlingen latere leerstofonderdelen met succes konden voltooien. In tabel 6.3. zijn drie cesuren of aftestgrensen vermeld voor de formatieve toetsen uit experiment I. Voor de door ons gehanteerde 80%-cesuur zijn overeenkomstig het gestelde in par. 6.4. de Bayes- en de maximin-utiliteit berekend*. De Bayes-utiliteit is in

*Hierbij is gebruik gemaakt van een op het Cito te Arnhem aanwezig experimenteel computerprogramma van N.H. Veldhuijzen.

Tabel 6.3. Aftestgrenzen voor de formatieve toetsen experiment I.

Formatieve toets/ subtoets	Aantal items	80% cesuur	Bayes- utiliteit	80%	maximin- utiliteit	80%	Bayes- cesuur	Bayes- utiliteit	maximin- cesuur	maximin- utiliteit
● Toets 1	30	24	.802		.677		24	.802	25	.722
subtoets 1.1	10	8	.710		.530		8	.710	9	.606
subtoets 1.2	10	8	.587		.434		9	.606	9	.476
subtoets 1.3	10	8	.816		.461		7	.848	9	.679
● Toets 2	24	19	.667		.425		16	.714	20	.590
subtoets 2.1/2.2	12	10	.863		.608		8	.910	11	.720
subtoets 2.3/2.4	12	10	.667		.602		12	.952	10	.602
● Toets 3	22	18	.867		.711		17	.880	19	.807
subtoets 3.1	14	11	.865		.478		10	.880	12	.706
subtoets 3.2	8	6	.834		.409		6	.834	7	.690
● Toets 4	22	18	.808		.796		18	.808	18	.796
subtoets 4.1	12	10	.725		.694		10	.725	10	.694
subtoets 4.2	10	8	.734		.646		9	.758	8	.646
● Toets 5	78	62	*		*		*	*	*	*
subtoets 5.1	50	40	.861		.799		40	.861	41	.849
subtoets 5.2	12	10	.812		.547		7	.921	11	.603
subtoets 5.3	16	13	.896		.724		13	.896	14	.860
● Toets 6	20	16	**		**		**	**	**	**
subtoets 6.1	6	5	.826		.584		5	.826	6	.647
subtoets 6.2	14	11			**		**	**	**	**
● Toets 7	12	10	.986		.537		5	.999	11	.791

* Niet berekend i.v.m. beperkingen computerprogramma.

** Niet berekend vanwege te grote numerieke onnauwkeurigheid.

dit geval berekend volgens een empirisch-Bayesiaanse methode, waarbij de 'a priori'-verdeling van θ geschat wordt uit de data.

Daarnaast is zowel de Bayes- als de maximin-cesuur bepaald waarvoor de Bayes- resp. maximin-utiliteit het hoogst is. De Bayes-utiliteit betreft hier de kans op een correcte toewijzing, terwijl de maximin-utiliteit de ondergrens van de kans op een correcte toewijzing betreft. Bij alle berekeningen zijn de utiliteiten u_{11} en u_{22} gelijk aan '1' en de utiliteiten u_{12} en u_{21} gelijk aan '0' gesteld. Met andere woorden aan de beslissing om een leerling te laten slagen indien zijn score op de latente variabele θ hoger of gelijk is dan de grensscore θ_0 , dan wel aan de beslissing om een leerling te laten zakken indien zijn score op de latente variabele θ lager is dan de grensscore θ_0 wordt een gelijke utiliteit toegekend.

Bij de vaststelling van de aftestgrenzen middels de in tabel 6.3. aangegeven drie methoden moet rekening gehouden worden met het feit dat zowel de Bayes- als de maximin-procedure de cesuur bepalen m.b.t. de latente variabele θ , en op basis daarvan op de gebruikte toets. De ene keer kan dat dus resulteren in een hogere of lagere geobserveerde grensscore op de toets dan de andere keer. Ook als deze toetsen geacht worden betrekking te hebben op hetzelfde itemdomein. Steeds wordt die grensscore gekozen waarvoor de Bayes- resp. maximin-utiliteit maximaal is. Geheel anders echter ligt dit bij de zgn. 80%-methode. Ook hier gaat het er in feite om dat een geslaagde leerling niet alleen een 80%-score kan behalen op deze toets, maar in principe op elke willekeurige toets m.b.t. het leerstofdomein waarin men geïnteresseerd is. Echter men neemt in het algemeen zonder meer aan dat een 80%-score op het domein, ook betekent een 80%-score op de toets. Er wordt derhalve - ten onrechte - geen rekening gehouden met toevallige kenmerken van deze specifieke toets. Vanuit theoretisch gezichtspunt zijn de Bayes- en de maximin-procedure veruit superieur t.o.v. de 80%-procedure. Anderzijds is de laatste intuïtief aantrekkelijk én rekenkundig ongecompliceerd. Beschouwen we nu de empirische gegevens in tabel 6.3, dan kunnen we constateren dat in 90% van de cesuurbepalingen de 80%-cesuur valt tussen, of gelijk is aan die van de twee andere methoden. Slechts 1 maal, bij subtoets 1.2 is de 80%-cesuur lager dan die bij de beide andere methoden. Bij deze subtoets zijn ook de bijbehorende utiliteiten het laagst in vergelijking met de andere toetsen.

Gegeven de data kunnen we stellen dat de gevonden utiliteiten bij de 80%-cesuurmethode geen aanleiding geven tot ernstige

zorg. Een uitzondering wordt wellicht gevormd door wederom subtoets 1.2. Echter de hoogste Bayes- en maximin-utiliteiten zijn niet veel hoger dan die behorend bij de 80% cesuur-methode. Kennelijk is deze subtoets minder geschikt voor het doel waarvoor hij is ontworpen. Tenslotte zij nog opgemerkt dat voor Toets 5 (totaal), toets 6 (totaal) en subtoets 6.2 geen utiliteiten berekend konden worden vanwege bepaalde beperkingen van het computerprogramma.

6.6.3. Consistentiebepaling formatieve toetsen experiment I

In tabel 6.2 is voor alle formatieve toetsen en subtoetsen de klassieke schatting van de ondergrens der betrouwbaarheid gegeven. De KR-20 betrouwbaarheidscoëfficiënt voldoet echter niet zo goed voor criterium-georiënteerde toetsen i.v.m. de relatief geringe lengte van de toetsen en de geringe spreiding. Bovendien is de spreiding voor criterium-georiënteerde toetsen niet relevant.

Zoals reeds eerder betoogd, verdient het de voorkeur te kijken naar de betrouwbaarheid van classificatiebeslissingen m.b.t. de leerstofbeheersing i.p.v. naar de betrouwbaarheid van de toets. In het navolgende zullen wij ons tot de eerstgenoemde beperken en die aanduiden met consistentie. Bij de berekening van de consistentie-coëfficiënten hebben wij twee methoden gehanteerd, enerzijds de methode van Huynh en anderzijds de methode van Subkoviak. Deze coëfficiënten zijn bepaald voor drie aftestgrenzen, t.w. de 80%-cesuur, de Bayes-cesuur en de maximin-cesuur. De resultaten van de berekeningen zijn weergegeven in tabel 6.4.

Voor de toetsen 5, 6 en 6.2 zijn voor de Bayes- en maximin-cesuur geen coëfficiënten berekend daar deze cesuren niet bepaald konden worden. Het is interessant om in tabel 6.4 de Huynh-coëfficiënten voor de drie aftestgrenzen met elkaar te vergelijken, hetzelfde voor de drie Subkoviak-coëfficiënten en voorts de Huynh- en Subkoviak-coëfficiënten met elkaar te vergelijken.

De overeenstemming tussen de Huynh-coëfficiënten voor de drie aftestgrenzen is relatief hoog: de correlaties tussen de kolommen 1, 3 en 5 zijn resp. .99, .99 en .98. De overeenstemming tussen de Subkoviak-coëfficiënten is daarentegen veel geringer voor de diverse aftestgrenzen. De correlaties tussen de kolommen 2, 4 en 6 zijn resp. .39, .89 en .15. Met name de correlatie tussen kolom 4 en 6, zijnde .15, valt op.

De beoordeling van de grootte van de Huynh- (κ) en Subkoviak-coëfficiënt (p_0), alsmede de onderlinge vergelijking kan

Tabel 6.4. Consistentie-coëfficiënten voor de formatieve toetsen experiment I bij drie aftestgrenzen.

Formatieve toets/ subtoets	80%-cesuur		Bayes-cesuur		maximin-cesuur	
	Huynh	Subkoviak	Huynh	Subkoviak	Huynh	Subkoviak
• Toets 1	.41	.74	.41	.74	.41	.71
subtoets 1.1	.23	.67	.23	.67	.22	.62
subtoets 1.2	.07	.58	.06	.56	.06	.56
subtoets 1.3	.17	.77	.13	.90	.19	.62
• Toets 2	.06	.61	.02	.94	.06	.53
subtoets 2.1/2.2	.27	.81	.17	.96	.29	.69
subtoets 2.3/2.4	.05	.57	.01	.95	.05	.57
• Toets 3	.45	.81	.42	.86	.46	.77
subtoets 3.1	.25	.83	.20	.91	.28	.72
subtoets 3.2	.28	.84	.28	.84	.32	.72
• Toets 4	.48	.74	.48	.74	.48	.74
subtoets 4.1	.28	.64	.28	.64	.28	.64
subtoets 4.2	.36	.68	.33	.69	.36	.68
• Toets 5	.61	.84	*	*	*	*
subtoets 5.1	.60	.81	.60	.81	.60	.80
subtoets 5.2	.06	.64	.01	.99	.06	.53
subtoets 5.3	.55	.86	.55	.86	.56	.82
• Toets 6	.62	.90	*	*	*	*
subtoets 6.1	.35	.78	.35	.78	.34	.67
subtoets 6.2	.56	.91	*	*	*	*
• Toets 7	.13	.98	**	1.00	.20	.92

* Zie opmerking tabel 6.3

** Coëfficiënt kappa is onbepaald, daar $p=q=1$.

niet zonder meer geschieden.

Bij coëfficiënt kappa, zowel als bij de Subkoviak-coëfficiënt, simuleren we in feite nog zo'n toets als waarmee we gewerkt hebben, met dezelfde marginalen. Tussen de beslissingen die we op grond van deze twee toetsen nemen kan overeenstemming bestaan op grond van toeval. Coëfficiënt kappa geeft nu de mate van overeenstemming aan boven die op grond van toeval verwacht kan worden.

De coëfficiënt van Subkoviak daarentegen geeft de mate van overeenstemming weer tussen de beslissingen op grond van de gebruikte toets en op grond van de gesimuleerde (fictieve) toets. In hoeverre deze overeenstemming mede een gevolg is van toeval doet daarbij niet ter zake. De maximale en de minimale grootte van de Subkoviak-coëfficiënt zijn afhankelijk van de keuze van de cesuur. De coëfficiënt bereikt een absoluut minimum als de cesuur samenvalt met de mediaan. Fig. 6.4 brengt de onderlinge relatie tussen de Huynh- (κ) en Subkoviak-coëfficiënt (p_0) in beeld, alsmede de relatie tussen coëfficiënten en aftestgrenzen.

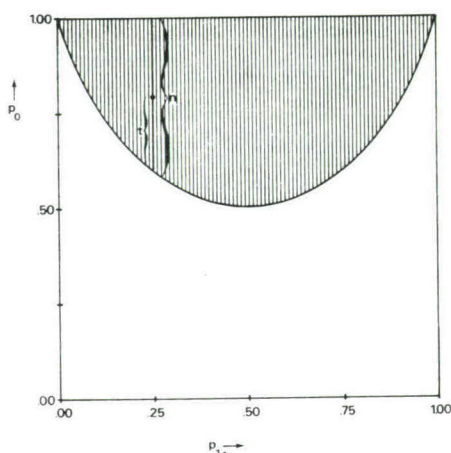


Fig. 6.4 Relatie tussen de consistentie-coëfficiënten

In fig. 6.4 zijn langs de horizontale as de waarden van de marginaal $p_{1.} = p_{.1}$ afgezet en langs de verticale as de waarden van de coëfficiënt van Subkoviak (p_0). Aangetoond kan worden m.b.v. de matrix en de bijbehorende marginalen (zie fig. 6.2) dat bij onafhankelijkheid van de toetsen de minimale waarde van de Subkoviak-coëfficiënt gelijk is aan

$$(\min)p_0 = 2 \left\{ \left(p_1 - \frac{1}{2} \right)^2 + \frac{1}{4} \right\} = p_e.$$

Coëfficiënt p_0 bereikt een minimum voor $p_1 = \frac{1}{2}$. De grafische voorstelling van de functie is een parabool. Coëfficiënt κ kan ook afgebeeld worden in fig. 6.4. Immers er geldt:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} = \frac{t}{n}$$

De geobserveerde Subkoviak-coëfficiënt p_0 ligt in fig. 6.4 in het gearceerde gedeelte, waarbij de afstand tot de parabool aangeeft hoeveel overeenstemming er meer is dan op grond van toeval verwacht kan worden. Deze afstand, uitgedrukt als proportie van de maximale mogelijke afstand, is κ .

6.7. Psychometrische analyse formatieve toetsen experiment II

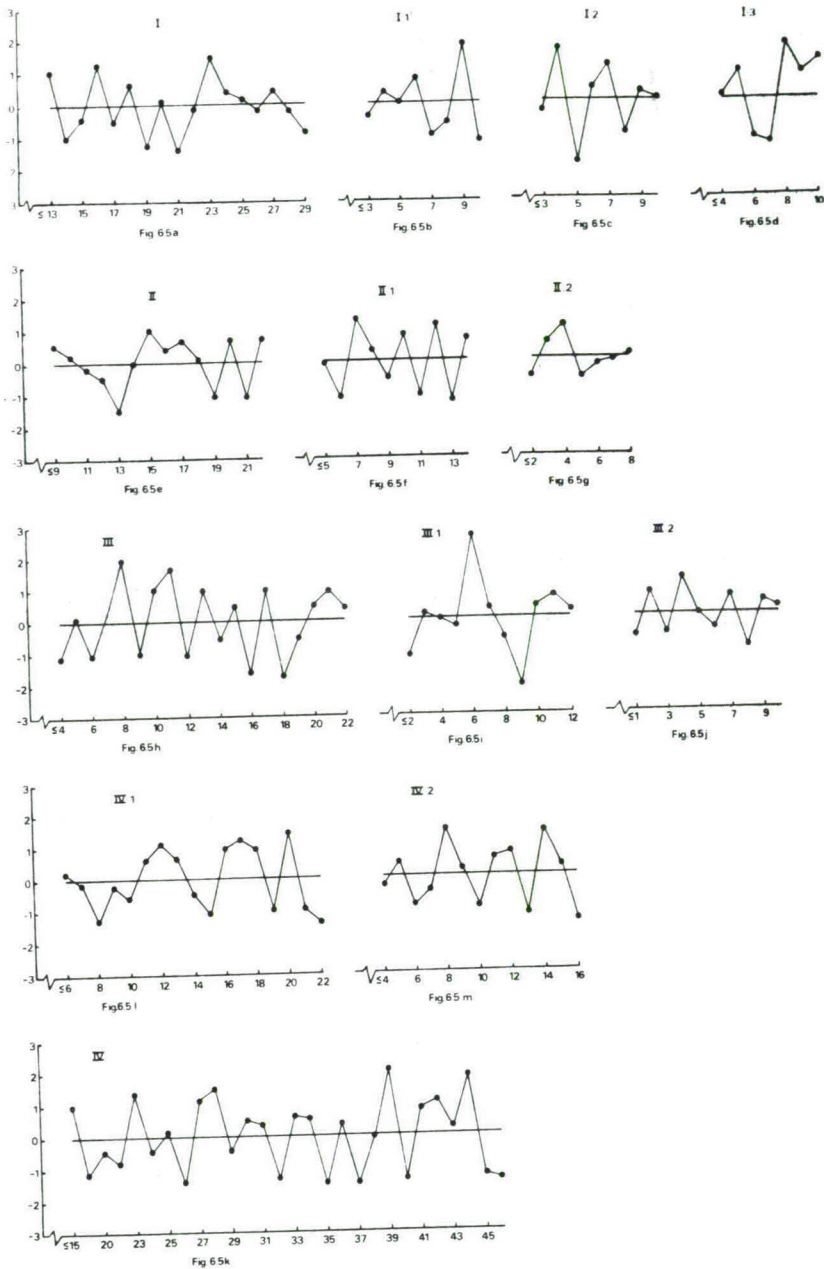
Analoog aan de in de voorgaande paragraaf beschreven cesuur- en consistentiebepaling van de toetsen uit experiment I zijn dezelfde procedures toegepast op de formatieve toetsen uit experiment II.

6.7.1. Modeltoetsing

Met behulp van de χ^2 -toets werd de assumptie dat de toetscores een negatief-hypergeometrische verdeling volgen getoetst. Tabel 6.5 bevat de resultaten van deze significantietoetsing.

Tabel 6.5 Modeltoetsing formatieve toetsen experiment II.

Formatieve toets/ df subtoets	χ^2	Formatieve toets/ df subtoets	χ^2
• Toets 1	14	• Toets 3	16
subtoets 1.1	5	subtoets 3.1	8
subtoets 1.2	5	subtoets 3.2	7
subtoets 1.3	4		
• Toets 2	11	• Toets 4	25
subtoets 2.1	7	subtoets 4.1	14
subtoets 2.2	4	subtoets 4.2	7
		subtoets 4.3	10



Figuur 6.5. Geconstateerde afwijkingen van de toetsscores (Exp. II)

Voor geen der toetsen bleek de geobserveerde verdeling significant af te wijken van de theoretisch verwachte negatief-hypergeometrische verdeling.

In fig. 6.5. (a t/m m) is wederom voor alle toetsen in beeld gebracht in welke mate de geobserveerde toetsscores afwijken van de theoretisch verwachte scores.

Evenals bij de formatieve toetsen uit experiment I kan ook nu geconcludeerd worden, dat het patroon van afwijkingen zoals dat uit de grafieken naar voren komt niet significant afwijkt van een aselekt karakter. Bij de bepaling van de aftestgrenzen en de consistentiecoëfficiënten kan derhalve ervan uitgegaan worden dat aan de assumptie met betrekking tot de negatief-hypergeometrische verdeling is voldaan.

6.7.2. Cesuurbepaling formatieve toetsen experiment II

In tabel 6.6 wordt een overzicht gegeven van enige beschrijvende gegevens betreffende de in het tweede experiment gebruikte formatieve toetsen en subtoetsen.

Tabel 6.6 Descriptieve gegevens formatieve toetsen experiment II

Formatieve toets/ subtoets	N	k	M	SD	KR-20	SE
• Toets 1	78	30	22.31	3.86	.69	2.15
subtoets 1.1	78	10	7.08	1.79	.48	1.29
subtoets 1.2	78	10	7.22	1.61	.38	1.27
subtoets 1.3	78	10	8.01	1.52	.43	1.15
• Toets 2	77	22	18.29	3.25	.76	1.58
subtoets 2.1	77	14	11.55	2.17	.66	1.26
subtoets 2.2	77	8	6.74	1.56	.65	.92
• Toets 3	77	22	15.09	4.49	.83	1.88
subtoets 3.1	77	12	8.42	2.50	.71	1.35
subtoets 3.2	77	10	6.68	2.41	.71	1.29
• Toets 4	78	48	33.92	7.24	.87	2.57
subtoets 4.1	78	22	15.32	3.91	.81	1.72
subtoets 4.2	78	10	7.62	2.33	.77	1.13
subtoets 4.3	78	16	10.99	2.78	.74	1.42

In de tweede kolom is het aantal proefpersonen vermeld. De

Tabel 6.7. Aftestgrenzen voor de formatieve toetsen experiment II

Formatieve toets/ subtoets	Aantal items	80% cesuur	Bayes- utiliteit 80%	maximin- utiliteit 80%	Bayes- cesuur	Bayes- utiliteit	maximin- cesuur	maximin- utiliteit
• Toets 1	30	24	.793	.777	25	.815	24	.777
subtoets 1.1	10	8	.691	.658	10	.801	8	.658
subtoets 1.2	10	8	.630	.600	10	.827	8	.600
subtoets 1.3	10	8	.709	.541	8	.709	9	.607
• Toets 2	22	18	.856	.784	18	.856	19	.792
subtoets 2.1	14	11	.814	.620	11	.814	12	.800
subtoets 2.2	8	6	.835	.541	7	.837	7	.775
• Toets 3	22	18	.865	.840	19	.865	18	.840
subtoets 3.1	12	10	.810	.804	11	.812	10	.804
subtoets 3.2	10	8	.803	.770	9	.836	8	.770
• Toets 4	48	38	.866	.858	40	.883	38	.858
subtoets 4.1	22	18	.842	.806	19	.851	18	.806
subtoets 4.2	10	8	.834	.739	8	.834	9	.774
subtoets 4.3	16	13	.798	.790	14	.839	13	.790

gegevens van één proefpersoon konden bij de toetsen 2 en 3 niet verwerkt worden. Kolom 3 bevat het aantal items per toets/subtoets, de volgende kolommen vermelden resp. het gemiddelde, de standaardafwijking, de homogeniteit en de standaardmeetfout.

Opvallend zijn in tabel 6.6 de relatief hoge homogeniteits-coëfficiënten, uitgezonderd die voor de drie subtoetsen van de eerste formatieve toets.

Tabel 6.7 bevat de diverse aftestgrenzen voor de formatieve toetsen.

Uit de gegevens in tabel 6.7 kan men opmaken dat in alle gevallen de 80% cesuur gelijk of lager is dan zowel de Bayes- als de maximin-cesuur. Opvallend is voorts dat de maximin-cesuur voor toets 2 en de bijbehorende subtoetsen hoger is dan zowel de 80%, als de Bayes-cesuur.

Geconstateerd kan worden dat bij hantering van de 80%-cesuur de leerlingen eerder als geslaagd beschouwd worden dan bij de beide andere methoden van cesuurbepaling.

6.7.3. Consistentiebepaling formatieve toetsen experiment II

Evenals voor de formatieve toetsen uit experiment I is ook voor de toetsen in het tweede experiment een consistentiebepaling uitgevoerd. De consistentie-coëfficiënten zijn berekend door middel van de methode van Huynh en de methode van Subkoviak. Voor elk der drie gehanteerde aftestgrenzen, t.w. de 80%-cesuur, de Bayes-cesuur en de maximin-cesuur, zijn deze coëfficiënten bepaald. De resultaten van deze berekeningen zijn samengevat in tabel 6.8. Uit tabel 6.8 valt af te lezen dat in vrijwel alle gevallen (uitgezonderd subtoets 2.2) de Huynh-coëfficiënten voor de 80%-cesuur hoger zijn dan de vergelijkbare coëfficiënten voor de Bayes- en de Maximin-cesuur. Voor de Subkoviak-coëfficiënt is dit beeld meer wisselend. Grote discrepanties doen zich echter niet voor.

Voor een vergelijking van de overeenkomsten tussen de diverse coëfficiënten zijn weer correlaties berekend. Deze zijn vermeld in tabel 6.9.

De correlaties tussen de Huynh-coëfficiënten zijn alle zeer hoog en vertonen het beeld dat ook reeds in experiment I naar voren kwam. Ook de correlaties tussen de Subkoviak-coëfficiënten vertonen het uit experiment I bekend beeld. Vooral de Subkoviak-coëfficiënt voor de Bayes-cesuur is - eveneens overeenkomstig de trend van experiment I - een uitbijter in het gezelschap.

Tabel 6.8 Consistentie-coëfficiënten voor de formatieve toetsen experiment II bij drie aftestgrenzen

Formatieve toets/sub-toets	80% cesuur		Bayes cesuur		Maximin cesuur	
	Huynh	Subkoviak	Huynh	Subkoviak	Huynh	Subkoviak
• Toets 1	.44	.73	.42	.75	.44	.73
subtoets 1.1	.27	.64	.13	.88	.27	.64
subtoets 1.2	.17	.59	.07	.89	.17	.59
subtoets 1.3	.24	.66	.24	.66	.23	.62
• Toets 2	.55	.80	.55	.80	.55	.78
subtoets 2.1	.44	.78	.44	.78	.44	.73
subtoets 2.2	.47	.84	.49	.77	.49	.77
• Toets 3	.59	.81	.56	.83	.59	.81
subtoets 3.1	.46	.74	.41	.79	.46	.74
subtoets 3.2	.50	.76	.46	.79	.50	.76
• Toets 4	.61	.82	.59	.85	.61	.82
subtoets 4.1	.51	.78	.48	.81	.51	.78
subtoets 4.2	.57	.79	.57	.79	.55	.78
subtoets 4.3	.39	.73	.35	.79	.39	.73

Over het geheel genomen kan geconcludeerd worden dat de consistentie-coëfficiënten voor de toetsen en subtoetsen van voldoende grootte zijn. De 80%-cesuur misstaat in het geheel niet tussen de beide andere - theoretisch onderbouwde - methoden. Of dit echter ook in andere situaties, bij andere toetsen, proefpersonen en leerstof opgaat, blijft echter onbekend.

Tabel 6.9 Correlatiematrix van consistentie-coëfficiënten experiment II.

Consistentie-coëfficiënt	H80	S80	HB	SB	Hm	Sm
Huynh / 80% (H80)	1.00	.92	.98	.01	.99	.99
Subkoviak / 80% (S80)		1.00	.96	-.14	.94	.95
Huynh / Bayes (HB)			1.00	-.17	.97	.97
Subkoviak / Bayes (SB)				1.00	.02	.02
Huynh / maximin (Hm)					1.00	.99
Subkoviak / maximin (Sm)						1.00

EXPERIMENT I: TOETSING VAN HET MODEL

7.1. Inleiding

In dit hoofdstuk worden de opzet, de uitvoering en de resultaten van een onderzoek beschreven gericht op een systematische toetsing van het door Bloom voorgestelde model betreffende het schoolse leren. Het model en de algemene vraagstellingen zijn achtereenvolgens beschreven in de hoofdstukken 3 en 4 van dit proefschrift. Naast de toetsing van het model was het onderzoek ook gericht op kennisverwerving aangaande de aard en vorm van toe te passen feedback en correctieve procedures in onderwijsleersituaties. In het onderzoek wordt een mastery- en een non-mastery-conditie onderscheiden. De mastery-conditie wordt verder onderscheiden in een Bloom- en een Keller-type, terwijl de non-mastery-conditie bestaat uit twee groepen waarbij de ene wel en de andere geen feedback krijgt. In totaal onderscheiden we dus vier condities. In het navolgende zullen wij aandacht besteden aan de hypothesen betreffende de toetsing van het model, en aan die betreffende de aard van de feedback.

7.1.1. Vraagstellingen betreffende het model

Een cruciaal element in het model van Bloom is de veronderstelling dat leerprestaties, na beëindiging van een leertaak, door feedback en correctieve procedures nog veranderd kunnen worden. In een recent artikel van Bloom (1980) ziet hij als de nieuwe richting waarin het onderzoek zich moet begeven: het zoeken naar veranderbare variabelen in het onderwijsleerproces. In het door ons uitgevoerde experiment verwachten we dan ook dat de leerprestaties na feedback en correctie hoger zullen zijn dan ervoor. Eveneens verwachten we dat de correlatie tussen leerprestaties na correctie en leerprestaties in de volgende leertaak vóór correctie afneemt in vergelijking met de correlatie met ongecorrigeerde leerprestaties in de voorafgaande leertaak. Tenslotte verwachten we dat de variantie in leerprestaties in de groep die feedback en correctie ontvangt in de opeenvolgende leertaken afneemt in vergelijking met de groep die geen feedback en correctie ontvangt.

De tot nu toe geformuleerde vraagstellingen hebben betrekking op leerprestaties die tijdens het onderwijsleerproces gemeten

worden. Indien een en ander niet empirisch aangetoond kan worden, betekent dit dat het model in werkelijkheid niet functioneert zoals dat in theorie zou moeten. Dit impliceert dan tevens dat effecten op summatieve toetsen zoals geclaimd door het model, in ons geval dan ook niet uitsluitend aan de werking van het model toegeschreven kunnen worden. Het model-zoals-gerealiseerd functioneert dan immers niet als het model-zoals-bedoeld. De operationalisatie van het door Bloom gepostuleerde model is dan invalide. In het door ons uitgevoerde experiment zullen we dan ook eerst nagaan of het model wel 'werkt' zoals verwacht, vóórdat de verklarende waarde van het model met betrekking tot leerprestaties wordt nagegaan.

7.1.1.1. Hoofdeffecten

a. Voorspellende waarde van het cognitief begingedrag

Bloom beredeneert dat het cognitief begingedrag (CBG) 50% van de variantie in leerprestaties op een summatieve toets verklaart. Hij maakt daarbij geen onderscheid tussen leerprestaties die na onderwijsleerprocessen met en zonder feedback en correctie tot stand zijn gekomen. Het lijkt redelijk te veronderstellen dat de invloed van het CBG in kwalitatief hoogstaande onderwijsleersituaties geringer is dan in kwalitatief mindere. In welke orde van grootte dit zal liggen, is voorspelbaar. We spreken de verwachting uit dat het percentage verklaarde variantie op een summatieve toets in een onderwijsleerproces waarin feedback en correctieve procedures opgenomen zijn, lager zal zijn dan 50%, terwijl dit percentage hoger dan 50% is in een onderwijsleerproces zonder feedback en correctie. Optimalisering van onderwijsleerprocessen zal waarschijnlijk leiden tot een geringere invloed van voorafgaand leren en -tot op zekere hoogte - van leerlingkenmerken op summatieve leerprestaties.

b. Voorspellende waarde van de affectieve leerlingkenmerken

De affectieve leerlingkenmerken verklaren conform het Bloom-model ongeveer 25% van de variantie in leerprestaties op een summatieve toets. Het is niet zonder meer inzichtelijk dat een optimaal onderwijsleerproces op de korte termijn tot een positievere attitude leidt dan een niet optimaal onderwijsleerproces. We verwachten dan ook geen verschil tussen condities waarin wel en condities

waarin geen feedback en correctieve procedures opgenomen zijn in het onderwijsleerproces. Dit impliceert derhalve dat het percentage verklaarde variantie, ongeacht de vormgeving van het onderwijsleerproces, ongeveer 25% zal bedragen.

c. Voorspellende waarde van de leertaak

Ook de leertaak - 'kwaliteit van de instructie' - verklaart volgens Bloom plm. 25% van de variantie in leerprestaties op een summatieve toets. Wederom kan verondersteld worden dat een optimaal vormgegeven leertaak een groter percentage variantie kan verklaren dan een kwalitatief minder instructieproces. We voorspellen derhalve dat een onderwijsleerproces waarin feedback en correctieve procedures opgenomen zijn meer dan 25% van de variantie in leerprestaties kan verklaren. Een onderwijsleerproces zonder feedback en correctie verklaart daarentegen minder dan 25%.

7.1.1.2. *Interacties*

Bloom veronderstelt dat het cognitief begingedrag en de affectieve leerlingkenmerken tezamen, dus inclusief de interactie van deze twee variabelen, plm. 65% van de variantie in leerprestaties verklaart. Wij veronderstellen wederom dat opname van feedback en correctieve procedures in een onderwijsleerproces ertoe leidt dat dit percentage lager zal zijn dan in het geval deze aspecten in een onderwijsleerproces ontbreken.

Het totale model tenslotte, inclusief alle interacties, zal ongeveer 90% verklaren. Dit percentage zal niet verschillen ongeacht de af- of aanwezigheid van feedback en correctieve procedures.

7.1.2. *De voorspellende waarde van het model na correctie voor tijd*

In het door Bloom voorgestelde model wordt geen rekening gehouden met de extra leertijd die feedback en correctieve procedures vergen. Weliswaar stelt Bloom dat de spreiding in benodigde leertijd tussen leerlingen in een mastery-systeem afneemt, maar onduidelijk blijft of de absolute leertijd ook zodanig gereduceerd wordt, dat deze geen concurrerende verklaring kan zijn voor eventuele hogere leerprestaties. Wij

veronderstellen dat er geen verschil in percentages verklaarde variantie is tussen meer en minder goed gestructureerde onderwijsleerprocessen, als er gecorrigeerd wordt voor verschillen in leertijd. De eerder geformuleerde hypothesen zullen dan ook tevens getoetst worden met in acht-neming van verschillen in leertijd.

7.1.3. Vraagstellingen betreffende feedback en correctieve procedures

In de hoofdstukken 3 en 4 is reeds nader ingegaan op diverse aspecten betreffende de aard en de vormgeving van feedback en correctieve procedures in onderwijsleerprocessen. In het hier beschreven experiment is een drietal hypothesen getoetst met betrekking tot feedback en correctiva.

In de eerste plaats is nagegaan of feedback en correctie ongeacht de wijze waarop, beter was dan geen feedback en correctie, waarbij zowel een toetsing plaatsvond zonder als een toetsing met correctie voor leertijd. De voorspelling is dat (ongeacht tijds-correctie) feedback en correctie beter is dan geen feedback.

In de tweede plaats is onderzocht of actieve feedback en correctie beter is dan passieve feedback en correctie. Ook hier is op twee manieren getoetst: met en zonder correctie voor leertijd. De voorspelling is dat zonder correctie voor leertijd actieve feedback beter is dan passieve feedback. Na correctie voor leertijd hebben we in feite geen argument om de ene methode superieur te achten aan de andere methode. We doen voor die situatie derhalve geen gerichte voorspelling.

In de derde en laatste plaats is tenslotte nagegaan of er verschillen in leerprestaties te onderkennen zijn bij de toepassing van totale versus partiële feedback (zie paragraaf 3.3.1.), eveneens met en zonder correctie voor verschillen in leertijd. Ook in dit geval hebben wij geen argumenten om de ene methode superieur te achten boven de andere. Geen gerichte voorspelling derhalve.

7.2. Proefpersonen

Aan het onderzoek namen in totaal 89 proefpersonen deel, 57 mannelijke en 32 vrouwelijke proefpersonen. De toewijzing aan de vier condities geschiedde aselekt met dien verstande dat - via gestratificeerde steekproeftrekking - de mannelijke, resp. vrouwelijke proefpersonen in gelijke mate over de condities verdeeld werden. In de loop van het onderzoek vielen om

diverse redenen (ziekte, verhuizing, e.d.) 5 proefpersonen af: 4 mannelijke en 1 vrouwelijke, verspreid over in totaal 3 condities.

In een eerste sessie werd ondermeer een vragenlijst aan de proefpersonen voorgelegd waarmee enige informatie betreffende (voor)opleiding, leeftijd, e.d. verzameld werd. Tabel 7.1 bevat een overzicht van deze gegevens.

Tabel 7.1. Overzicht van enige persoonsgegevens, der proefpersonen in experiment I (N=84)

		conditie			
		I	II	III	IV
● geslacht	mannelijk	16	12	13	12
	vrouwelijk	8	8	7	8
● leeftijd	gemiddelde	22.29	22.35	23.15	24.50
	range	19-29	19-27	19-37	19-39
	stand.dev.	2.76	2.80	4.36	5.84
● vooropleiding	MBO	-	-	1	1
	HAVO	1	-	-	3
	VWO	23	20	19	16
● huidige opleiding	HBO	2	-	1	4
	VWO	22	20	19	16
● aantal jaren Frans	gemiddelde	5.46	5.3	5.20	5.10
	range	3-8	4-7	3-8	3-9
	stand.dev.	1.41	1.22	2.09	1.45

Uit tabel 7.1 blijkt dat de gemiddelde leeftijd in de groepen III en IV wat hoger is dan die in de beide andere groepen. Dit wordt voornamelijk veroorzaakt doordat zich in de groepen III en IV een paar relatief oude deelnemers bevinden van resp. 32, 37, 38 en 39 jaar. Vooropleiding, alsmede de opleiding die gevolgd werd in de periode waarin het onderzoek plaatsvond, is voor de vier groepen vergelijkbaar. Wat betreft het aantal jaren dat de proefpersonen in het voorafgaand onderwijs les gehad hebben in de Franse taal kan geconstateerd worden dat het gemiddeld aantal jaren voor de verschillende groepen overeenkomt, evenals de spreiding. De tamelijk grote range

wordt veroorzaakt doordat niet alle proefpersonen in het voorafgaand onderwijs het vak Frans opgenomen hadden in hun vakkenpakket. Dit impliceert naar alle waarschijnlijkheid dat ook de individuele verschillen met betrekking tot de actieve en passieve beheersing van de Franse taal binnen de groepen aanmerkelijk zullen zijn.

De proefpersonen waren afkomstig uit het universitair en het hoger beroepsonderwijs. De eerste categorie werd gevormd door studenten van de Katholieke Hogeschool Tilburg, de tweede categorie voornamelijk door studenten van het Moller-instituut te Tilburg, een zgn. Nieuwe Lerarenopleiding (NLO). Deelname aan het onderzoek geschiedde op vrijwillige basis. Een deel der proefpersonen deed dit in het kader van de binnen de Subfaculteit Psychologie der K.H.T. gebruikelijke proefpersoonuren-regeling, een ander deel ontving een (geringe) financiële vergoeding. In beide gevallen was zowel het aantal uren, als het geldbedrag gefixeerd voor het gehele experiment. Het maakte voor de vergoeding derhalve niet uit of de proefpersonen veel of weinig tijd aan het onderzoek moesten besteden. Een en ander was van belang in verband met de tijdmetingen die in het experiment verricht werden.

Naast informatie betreffende allerlei formele aspecten zoals opleiding e.d. werd ook aandacht besteed aan meer informele zaken die van belang konden zijn m.b.t. aanvangskennis voor het experiment. Het is immers niet ondenkbaar dat diegenen die zich voor een experiment betreffende het leren van een tweede taal aanmelden, meer aan de studie van de desbetreffende taal gedaan hebben dan hetgeen gebruikelijk is binnen hun reguliere (voor)opleiding.

In tabel 7.2 zijn enige gegevens samengebracht aangaande handelingen gericht op verwerving van een actieve of passieve beheersing van de Franse taal.

De gegevens in tabel 7.2. overziend, is de conclusie gewettigd dat, afgezien van enige vakantievreugde, de verschillende mogelijke activiteiten nauwelijks enige bijdrage geleverd hebben aan de verhoging van de beheersing der Franse taal.

De proefpersonen verdeeld over de vier groepen zijn derhalve zowel qua formele als qua meer informele kenmerken tamelijk goed vergelijkbaar.

7.3. Instrumenten voor de meting der leerlingkenmerken

Conform het model van Bloom (zie: Hoofdstuk 3) zijn een aantal instrumenten gebruikt ter bepaling van de cognitieve en affectieve aanvangskenmerken. In de navolgende paragrafen worden deze instrumenten beschreven.

Tabel 7.2. Descriptieve gegevens betreffende extra studie van het Frans

Aard der activiteiten	Conditie							
	I		II		III		IV	
	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
Franse kranten e.d. lezen*	2.96	.20	3.00	.00	2.80	.41	2.90	.31
Luisteren/kijken Franse radio en t.v.*	2.54	.51	2.75	.44	2.70	.47	2.55	.51
Met Fransen spreken*	2.75	.44	2.65	.49	2.55	.61	2.60	.50
In het Frans corresponderen*	3.00	.00	2.95	.22	2.95	.22	2.90	.45
Deelname Franse culturele activiteiten*	3.00	.00	3.00	.00	2.90	.31	3.00	.00
Bezoek Franstalig land**	1.92	.50	2.20	.77	1.80	.52	1.95	.39

Scoring vond plaats op een driepuntsschaal met als schaalpunten:

* (1) dikwijls, (2) soms, (3) zelden.

** meer dan 1 maand, (2) 1 maand of minder, (3) nooit.

7.3.1. Instrumenten ten behoeve van het cognitieve begin-gedrag

• Leerdoelgebonden voortoets

Op analoge wijze als bij de in hoofdstuk 5 beschreven constructie van de formatieve toetsen werd een leerdoelgebonden voortoets samengesteld. Elke doelstelling van elke onderwijsleereenheid of leertaak werd door tenminste één item in de toets gerepresenteerd. De vormgeving van de items, alsmede de correspondentie tussen item en doelstelling werd weer beoordeeld door een vakdeskundige van het Talencentrum der K.H.T. Een en ander vormde een garantie voor de inhoudsvaliditeit van de toets. Het totaal

aantal items in de toets bedroeg 138. In verband met de omvang van de toets werd deze in twee sessies afgenomen: de eerste maal 49 items en de tweede maal de resterende 89 items.

● Modern Language Aptitude Test (MLAT)

De MLAT is ontwikkeld door J.B. Carroll en S.M. Sapon gedurende de jaren 1953-1958 tijdens een onderzoeksproject aan de Harvard University in de U.S.A.

De MLAT is een aanlegtest die een indicatie geeft van de waarschijnlijkheid waarmee een individu met succes een vreemde taal kan leren. De test kan gebruikt worden om de kans op succes in het leren spreken en begrijpen van een vreemde taal te prediceren, maar is evenzeer geschikt voor het voorspellen van de kans op succes ten aanzien van het leren lezen, schrijven en vertalen van een vreemde taal.

De door Carroll & Sapon (1959) gepresenteerde validiteitsgegevens maken duidelijk dat de test voldoende geschikt is voor de predictiedoeleinden waarvoor hij ontworpen is. Voor de predictie van succes met betrekking tot het leren van het Frans door universitaire studenten rapporteren Carroll & Sapon validiteitscoëfficiënten variërend van .27 tot .69, met een gemiddelde van .49. Voorts liggen de correlaties tussen de MLAT-scores met criteriumscores bij taalopleidingen ongeveer .20 hoger dan de correlaties tussen algemene intelligentiescores en deze criteria.

Bij de interpretaties van deze gegevens in het licht van het in dit hoofdstuk beschreven onderzoek moet overigens wel aangetekend worden, dat de validatie heeft plaatsgevonden bij individuen waarvan het Engels de moedertaal vormde.

De MLAT bestaat uit vijf subtests: (1) 'Number learning', deze component is van belang bij het auditief begrijpen van een vreemde taal; (2) 'Phonetic script', dit onderdeel meet de vaardigheid in het leren van de klank/symbool-correspondentie; (3) 'Spelling clues', meet ongeveer hetzelfde als de tweede subtest, maar is in hoge mate gebaseerd op kennis van de Engelse vocabulaire; (4) 'Words in sentences', deze subtest meet de vaardigheid met betrekking tot het omgaan met de grammaticale aspecten van een vreemde taal; (5) 'Paired associates', dit onderdeel meet het korte termijngeheugen-aspect met betrekking tot het leren van een vreemde taal.

In het kader van ons experiment is uitsluitend gebruik gemaakt van subtest 4 'Words in sentences' en subtest 5

'Paired associates'. Deze subtests doen slechts een gering beroep op kennis van de Engelse taal. Bovendien zijn ze in het kader van de leertaken in het experiment het meest relevant.

De MLAT vormt tezamen met de leerdoelgebonden voortoets in ons onderzoek de operationalisatie van het cognitief begingedrag.

7.3.2. Instrumenten met betrekking tot de affectieve aanvangskenmerken

In het model van Bloom draait het bij de affectieve leerlingkenmerken vooral om de attitudes ten aanzien van het vak, de school en de eigen persoon. In het onderzoek naar de invloed van attitude en motivatie op het succes bij het leren van een vreemde taal gaat het daarentegen heel vaak om attitudes ten aanzien van de leden van de (bv.) Franse taalgroep en de Franse cultuur. Dergelijke attitudes zijn namelijk bepalend voor de motivatie waarmee een individu een vreemde taal leert en bijgevolg voor een belangrijk deel verklarend voor zijn leerresultaten. Uit de literatuur (cf. Knibbeler, 1976) blijkt echter dat dergelijke attitudes vooral bij een langdurige talenstudie een succes bepalende factor vormen. Aangezien het in ons onderzoek een cursus van korte duur betreft, is toetsing van dit aspect van minder belang.

Geheel in lijn met de opvattingen van Bloom menen wij dat het in dit onderzoek meer voor de hand ligt de attitude van de proefpersonen ten aanzien van de vormgeving van de cursus te meten, alsmede de meer directe belangstelling voor het leerobject.

Als instrument voor de meting van de affectieve leerlingkenmerken is gebruik gemaakt van een aangepaste versie van een door Knibbeler (1976) geconstrueerde attitudeschaal.

Aansluiting wordt gezocht bij de omschrijving van Rosenberg (1967) van een attitude, als 'a relatively stable affective response to an object'. De meting van een attitude omvat dan twee aspecten: het bepalen van de richting (positief of negatief) en het bepalen van de mate van affect.

De door Fishbein en Rosenberg ontwikkelde methode levert, via de cognitieve structuur, een indirecte meting op van de attitude. Deze cognitieve structuur bestaat uit opvattingen over de mate waarin een bepaald object de realisering van datgene wat als waardevol gezien wordt, mogelijk maakt of belemmert. Meting van de cognitieve structuur betekent derhalve dat proefpersonen enerzijds een aantal met betrekking tot het attitude-object relevante waarden moeten evalueren, en

(een deel van het) onderwerp van deze studie.

7.4.1. De leertaak

Ten aanzien van de leertaak wordt in het model van Bloom een viertal aspecten onderscheiden (zie par. 3.3.), te weten 'cues, participation, reinforcement en feedback/correctives'. Deze aspecten bepalen in hoge mate de kwaliteit van de instructie.

'Cues'

De 'cues' in het lesmateriaal zoals gebruikt in dit onderzoek worden voornamelijk gevormd door de in de leertaken opgenomen opgaven. Het is een bekend gegeven, waarvoor voldoende empirische evidentie aanwezig is (zie bijvoorbeeld Frase, 1968, Boyd, 1973; Wolters, 1975; Kirschner, 1980), dat het toevoegen van vragen en opgaven aan schriftelijk cursusmateriaal leidt tot betere leerresultaten. Het gunstige effect wordt op twee manieren bereikt. Ten eerste richten vragen de aandacht van de leerling op bepaalde delen van de leerstof. Op deze wijze krijgt de leerling als het ware een selectie-instrument in handen: alleen wat gevraagd wordt, hoeft geleerd (=onthouden) te worden, de rest kan minder aandacht krijgen. Indien de vragen de onderwijsdoelen niet volledig representeren kan dit een averechts effect hebben. Bij de vormgeving van de cursus is dan ook zorgvuldig nagegaan of de vragen en opgaven wel voldoende representatief waren voor de cursusdoelen. De genoemde selecterende werking van vragen heeft tevens een generaliserend effect: zij bevatten aanwijzingen over het *soort* zaken dat geleerd moet worden (cf. Rothkopf, 1966 en Rothkopf & Bisbicos, 1967).

De tweede manier waarop vragen en opgaven het leerresultaat kunnen beïnvloeden hangt nauw samen met de eerste: vragen en opgaven bevatten indicaties over welk soort leergedragingen moeten worden vertoond. Zij duiden met andere woorden aan welke vaardigheden bij kennisneming van de leerstof moeten worden beoefend. Indirect sturen de vragen derhalve het leergedrag van de leerling.

Buiten de in de lessen opgenomen vragen en opgaven geeft ook de leerdoelgebonden voorttoets (zie par. 7.3.1.) de leerlingen een idee welk soort gedrag vertoond moet worden. Ook de formatieve toetsen die aan het einde van elke leertaak of les afgenomen worden, hebben deze functie, althans voor die

leerlingen die niet het beheersingscriterium bereiken en nog aanvullende instructie behoeven. De toetsen fungeren in dit verband als een pre-instructionele strategie. Hartley & Davies (1976) bespreken in hun artikel een viertal van dergelijke pre-instructionele strategieën, t.w. voortoetsen, leerdoelen, overzichten en 'advance-organizers'. Zij noemen ondermeer een viertal factoren die op het effect van voortoetsen van invloed zijn:

1. De onderwijsperiode na de toets moet betrekkelijk kort zijn. Dit betekent dat de effecten van de voortoets groter zijn als deze betrekking heeft op een leerstofonderdeel van relatief beperkte omvang (bijv. een les) in vergelijking met een toets die de gehele cursus beslaat.
2. Het onderwijs mag niet 100% effectief zijn. Immers, als alle leerlingen de doelstellingen van het onderwijs bereiken en derhalve de natoets perfect maken, dan kan er geen verschil gevonden worden tussen de groep met en de groep zonder voortoets. In deze situatie treedt het zgn. plafondeffect op.
3. Voortoetsen functioneren beter naarmate de leerlingen ouder zijn (of intelligenter).
4. De leerlingen moeten al enige voorkennis betreffende het onderwerp van de voortoets en het onderwijs hebben. (Zonder enige voorkennis zullen zij anders zelfs de vragen niet eens begrijpen).

Gezien de vormgeving van de cursus, de frequente toetsing, alsmede de leeftijd en intelligentie van de aan het onderzoek meedoende proefpersonen lijken de algemene voortoets, de in de les opgenomen vragen en opgaven alsmede de formatieve toetsen voor die proefpersonen die het beheersingscriterium niet halen, te voldoen aan de door Hartley & Davies geformuleerde eisen.

'Participation'

De leerlingen die een leertaak moeten volbrengen, behoren iets te doen met de diverse 'cues', anders vindt er immers geen leren plaats. In het onderzoek werd er door de proefleider(s) op toe gezien dat alle proefpersonen aandacht besteedden aan de vragen en opgaven in de diverse leertaken. Op welke wijze zij dit deden en hoeveel tijd zij daaraan wilden besteden werd aan de proefpersonen zelf overgelaten. Een en ander maakte deel uit van het onderzoek met betrekking tot de bestede studietijd. Voorts werd uitgegaan van het principe dat de leertaken in het onderzoek zoveel mogelijk vergelijkbaar moesten zijn met die in gebruikelijke talencursussen, met

andere woorden gestreefd werd naar een zo groot mogelijke ecologisch valide onderzoeksopzet.

'Reinforcement'

De positieve of negatieve bevestiging ('reinforcement') speelt zonder meer een belangrijke rol bij het leren (cf. Hilgard & Bower, 1975). De reinforcement bij de leertaken in het kader van het hier beschreven experiment bestond *niet* uit de financiële vergoeding die gegeven werd. Deze was immers voor alle proefpersonen gelijk. De beloning voor een goede leerprestatie werd in feite gevormd door de door proefpersonen individueel behaalde tijdwinst.

Voorts kan verwacht worden dat de vrijwillig aan het onderzoek deelnemende proefpersonen, gezien hun intellectuele achtergrond het leveren van een goede leerprestatie ook als een positieve bevestiging zouden ervaren.

In het experiment en in de vormgeving van de leertaak is verder niet in meer specifieke zin aandacht besteed aan het reinforcement-aspect.

'Feedback en correctiva'

Binnen het model van Bloom betreffende het leren op school vormen terugkoppeling en correctieve procedures zeer belangrijke elementen. Het zijn dan ook deze factoren die in ons experiment gevarieerd worden.

Onderscheiden werden vier varianten van terugkoppeling en correctiva:

1. proefpersonen die op de aan het einde van elke les afgenomen formatieve toets (zie par. 5.2.4.) een score lager dan het 80%-criterium (zie par. 6.6.2.) behalen, berekend over alle toetsitems, moeten de gehele les opnieuw bestuderen ('*totale herhalingsconditie*');
2. proefpersonen die op de aan het einde van elke les afgenomen formatieve toets bij een of meer der subtoetsen (zie tabel 5.2.) een score lager dan het 80%-criterium behalen, berekend over de items van die subtoets, moeten het met die subtoets corresponderende lesonderdeel opnieuw bestuderen. In dat geval moet een verkorte, schematische versie van dat betreffende lesonderdeel bestudeerd worden ('*partiële herhalingsconditie*');
3. proefpersonen die op de aan het einde van elke les afgenomen formatieve toets een score lager dan het 80%-criterium behalen, berekend over alle toetsitems, krijgen

de toets terug, waarbij de gemaakte fouten verbeterd zijn. De proefpersonen dienen dan de toets met verbeteringen nogmaals door te nemen ('*kennis van correcte resultaten (KCR)-conditie*').

4. In de laatste variant tenslotte mochten de proefpersonen, ongeacht hun score op de eveneens steeds aan het einde van elke les afgenomen formatieve toets, verder gaan met de volgende les ('*controle conditie*').

Voor die proefpersonen uit de bovengenoemde condities 1 t/m 3 die de een of andere vorm van remediële instructie ontvangen hadden, volgde nog een tweede formatieve toets (zie par. 5.2.4.). Indien op deze toets als geheel (voor de 'totale herhalings'- en de 'KCR-conditie'), dan wel op een bepaalde subtoets (voor de 'partiële herhalingsconditie') de behaalde score lager was dan het 80%-criterium, dan volgde mondelinge uitleg door de proefleider. Dat wil zeggen aangegeven werd welke items fout beantwoord waren en wat het goede antwoord moest zijn. Tevens werd daarbij de grammaticale regel aangegeven waarop het goede antwoord gebaseerd was.

De correctieve procedure in de 'totale herhalingsconditie' (THC), waarin dus ook die lesonderdelen opnieuw bestudeerd moesten worden die wel in voldoende mate beheerst werden, komt het meest overeen met de werkwijze in de zgn. Keller-plan-benadering. De correctieve procedure in de 'partiële herhalingsconditie' (PHC) daarentegen is meer in overeenstemming met de benaderingen binnen Bloom's mastery-learning paradigma.

De condities 1 (THC) en 2 (PHC) vormen voorbeelden van *actieve* terugkoppeling in de zin dat de proefpersonen nog iets met de leerstof (de remediële instructie) moeten doen: opnieuw leerstof doornemen, opnieuw vragen en opgaven bekijken, een nieuw schematisch overzicht van de leerstof bestuderen, e.d. Conditie 3 (KCR) vormt een voorbeeld van een meer *passieve* terugkoppeling: de proefpersoon neemt kennis van de foute antwoorden plus de verbeteringen en daarbij blijft het. Hij hoeft dus in feite niets meer met de leerstof te doen. In de controle-conditie (conditie 4) tenslotte is in het geheel geen sprake van enigerlei vorm van feedback; noch actief, noch passief.

7.5. Instrumenten voor de meting der leerresultaten

In het model van Bloom komen tenslotte de metingen van de output aan de orde, namelijk die betreffende de (cognitieve) leerresultaten, de leertijd(en) en de affectieve resultaten van het onderwijsleerproces. In de navolgende paragrafen

worden de instrumenten, die in het experiment de operationalisatie vormden van deze output-metingen, besproken.

7.5.1. *Instrumenten ten behoeve van de cognitieve leerresultaten*

● Leerdoelgebonden tussentoetsen

Op analoge wijze als bij de in hoofdstuk 5 beschreven constructie van de formatieve toetsen werden twee leerdoelgebonden tussentoetsen samengesteld. De eerste tussentoets (TT-1) vormde een meting van de onderwijsdoelstellingen die in de leertaken 1 en 2 aan de orde kwamen, terwijl de tweede tussentoets (TT-2) de doelstellingen van de leertaken 3 en 4 bestreek. Elke doelstelling werd door minimaal één item in de toets gerepresenteerd. Screening der items door een vakdeskundige vormde wederom de basis van de inhoudsvaliditeit. Het aantal items in TT-1 bedroeg 27 en TT-2 telde 22 items.

● Leerdoelgebonden natoets

De leerdoelgebonden natoets was identiek aan de eerder afgenomen voortoets en bestond dus ook uit 138 items.

● Leerdoelgebonden retentietoets

De retentietoets was gelijk aan de natoets en de voortoets. Alleen het tijdstip waarop de toets werd afgenomen verschilde van dat der beide andere toetsen.

7.5.2. *Meting van de leertijd/leertempo*

Door de proefleider(s) werd voor alle proefpersonen door middel van een stopwatch de tijd bijgehouden die besteed werd door de proefpersonen aan

- het bestuderen van de leertaken
- het uitvoeren van de drie correctieve procedures
- het volgen van de eventueel benodigde mondelinge uitleg door de proefleider
- het maken van de leerdoelgebonden formatieve toetsen en formatieve hertoetsen (na een correctivum).

Op deze wijze kon voor alle proefpersonen de werkelijk bestede tijd bepaald worden aan het bestuderen van de leertaak, aan het maken van formatieve (her)toetsen en aan het uitvoeren van eventuele correctieve procedures.

7.5.2.1. Enige kanttekeningen bij het meten van studietijden

Het opnemen van formatieve (her)toetsen en correctieve procedures in onderwijsleerprocessen impliceert dat de benodigde studietijd voor een leerling bij een bepaalde leertaak toeneemt in vergelijking met leerlingen die geen toetsen en correctiva ontvangen. In ons onderzoek zou men dus op het eerste gezicht kunnen verwachten dat de proefpersonen in de condities 1 t/m 3 meer tijd consumeren dan die in conditie 4, terwijl de tijdsconsumptie in de condities 1 en 2 weer hoger zal zijn dan die in conditie 3. Met andere woorden (eventuele) hogere leerprestaties in een conditie in vergelijking met die in andere condities kunnen niet zonder meer toegeschreven worden aan een betere kwaliteit van de instructie. Misschien hadden de proefpersonen eenvoudigweg meer leertijd tot hun beschikking. Niet 'kwaliteit van de instructie' is dan de verklarende variabele, maar 'leertijd'. Faw & Waller (1976) onderscheiden een viertal methoden om studietijden te controleren of te manipuleren. Simons (1981) bespreekt deze methoden; hieraan worden de verschillende benamingen van de studietijdcontrolemethoden ontleend.

In de *neutrale vergelijkingsmethode* wordt de studietijd van de controle proefpersonen kunstmatig verlengd, doordat zij bezig worden gehouden met een neutrale taak, terwijl de proefpersonen in de experimentele conditie de manipulatie van de onafhankelijke variabele ondergaan.

In de tweede methode, de *ongelijke studietijdmethode*, krijgen de proefpersonen in de experimentele groep in vergelijking met die in de controlegroep extra tijd om bijv. toegevoegd leermateriaal (d.i. de onafhankelijke variabele) te bestuderen. Onder andere door Peeck (1970, 1977) is als bezwaar tegen deze methode aangevoerd dat het niet zonder meer mogelijk is om een eventueel gevonden effect toe te schrijven aan de manipulatie van de onafhankelijke variabele dan wel aan de extra studietijd.

Bij de derde methode, de zgn. *efficiëntiemethode*, krijgen de proefpersonen in de experimentele en de controlegroep allen evenveel studietijd. De proefpersonen in de experimentele conditie houden zich in deze tijd niet alleen bezig met de voor beide groepen gelijke leertaak, doch ook met de door de onderzoeker geïntroduceerde onafhankelijke variabele. Peeck (1970) heeft met name deze methode aanbevolen. Later onderzoek (Peeck, 1977) toonde echter aan dat de extra beschikbare studietijd niet altijd tot verhoging van de bestede studietijd leidde. Deze bevinding is vanuit de modellen van Carroll (par. 2.2.1.) en Bloom (hoofdstuk 3)

voorspel- en verklaarbaar.

Tenslotte is er dan nog een vierde methode, de *onbeperkte studietijdmethode*, waarbij in feite geen tijdslimiet wordt gesteld. Alle proefpersonen kunnen net zoveel tijd aan de leertaak (en de experimentele treatment) besteden als men zelf nodig acht. Deze methode heeft als voordeel dat zgn. efficiëntiescores kunnen worden berekend, waarbij een soort leerwinst per tijdseenheid kan worden berekend. Faw & Waller (1976) onderscheiden in dit verband:

- de absolute score: d.i. het groepsgemiddelde op de natoets;
- de efficiëntie-score: d.i. de ratio van de absolute score en de gemiddelde studietijd;
- de absolute index: d.i. de ratio van de absolute score van de experimentele en die van de controlegroep;
- de efficiëntie-index: d.i. de ratio van de efficiëntie-score van de experimentele en die van de controlegroep.

Bij de index > 1.00 geldt dan dat de experimentele groep beter is dan de controlegroep met betrekking tot leerprestaties en/of leerefficiëntie. Bij een index < 1.00 geldt het omgekeerde, of er is geen verschil. Simons (1981, p.133) noemt als bezwaar van de onbeperkte studietijdmethode vooral de onduidelijkheid die bestaat aangaande het stopcriterium dat proefpersonen hanteren. Afgezien van factoren als vermoeidheid en geen zin meer hebben, zijn er volgens Simons (o.c., p. 135) drie mogelijke criteria om eerder met het bestuderen van een leertaak op te houden dan in principe is toegestaan:

- a. *tijds criterium*: men vindt dat er voldoende tijd aan de leertaak besteed is;
- b. *herhalings criterium*: men vindt dat men de stof voldoende vaak herhaald heeft;
- c. *beheersings criterium*: men vindt dat men de leerstof in voldoende mate beheerst.

Buiten bovenstaande door Simons genoemde criteria kan men ook denken aan criteria van meer sociaal-psychologische aard: proefpersonen stoppen met hun leertaak als ze zien dat anderen, die gelijk of later begonnen zijn, er ook mee ophouden. Dergelijke processen waarin op de een of andere wijze een vorm van sociale controle kan optreden, kunnen zich met name in groepsgerichte vormen van onderwijs (en dat is het grootste gedeelte) voordoen. In de leertaken, waarbij leerlingen na een onvoldoende toetsresultaat, geconfronteerd worden met de een of andere correctieve procedure kan wellicht een vorm van 'trial-and-error'-gedrag optreden. De leerlingen bestuderen de leertaak slechts oppervlakkig: als de toets desondanks goed gemaakt wordt, is dat 'meegenomen'; wordt de toets daarentegen slecht gemaakt, dan is er nog geen man over boord: er is immers nog correctie mogelijk. Of, en in welke

mate dit gebeurt, kan worden nagegaan door het percentage goed beantwoorde items op de formatieve toets bij de eerste afname in de experimentele groep te vergelijken met dat in de controle-groep. Indien het percentage in de experimentele groep niet significant hoger is dan in de controle-groep, dan treedt dit 'trial-and-error'-gedrag kennelijk niet op.

Simons (o.c., p.138 e.v.) beveelt aan om bij de keuze van een studietijdcontrole methode enerzijds rekening te houden met de voor- en nadelen van de verschillende methoden, maar anderzijds ook met de aard van de onderzoeksvraag die men tracht te beantwoorden. Hij somt een vijftal vraagtypen op, waaronder ook onze onderzoeksvraag aangetroffen wordt. Deze kan immers als volgt geformuleerd worden: 'Hebben terugkoppeling en correctie naast een kwantitatief effect (via studietijdverlenging) ook een kwalitatief effect (via gunstige beïnvloeding van leerprocessen)?' Of: 'Zorgen feedback en correctiva voor efficiënter leren?'.

Uitgaande van deze onderzoeksvraag(-vragen) en de door Faw & Waller (1976) onderscheiden vier studietijdcontrole methoden komt vooral de onbepaalde studietijd methode in aanmerking. Het interpretatieprobleem dat dan nog resteert, betreft het stop-criterium: Op welke gronden stoppen de proefpersonen? Zijn deze in de verschillende groepen hetzelfde? Bij de beschrijving van de door ons gevolgde procedure bij opzet en uitvoering van het experiment zal aangegeven worden op welke wijze dit interpretatieprobleem wordt ondervangen.

7.5.3. Meting van de affectieve leerresultaten

Dezelfde attitudeschaal die is gebruikt voor de meting van de affectieve aanvangskenmerken der proefpersonen (zie par. 7.3.2.), is ook gebruikt voor de meting van de (eventuele) gevolgen van het onderwijsleerproces op de attitudes van de proefpersonen. Vergelijking van de waarden vóór en na het volgen van de cursus geeft een indruk van de eventuele attitudeveranderingen.

7.6. Procedure

In tabel 7.3. wordt een overzicht gegeven van de procedure zoals die in het eerste experiment tijdens de verschillende sessies voor de vier condities is gevolgd. Per week vonden steeds twee sessies plaats, waarbij minstens twee dagen tussen de sessies moesten zitten. Uitzondering hierop maakten de twee laatste sessies. Sessie 10 vond 1 week na sessie 9 plaats en

Tabel 7.3

Sessie	Conditie I	Conditie II	Conditie III	Conditie IV	Tijdregistratie
Sessie 1	- Introductie - Vragenlijst - Attitudeschaal - Voortoets-dl. 1	- Introductie - Vragenlijst - Attitudeschaal - Voortoets-dl. 1	- Introductie - Vragenlijst - Attitudeschaal - Voortoets-dl. 1	- Introductie - Vragenlijst - Attitudeschaal - Voortoets-dl. 1	
Sessie 2	- MLAT - Voortoets-dl. 2	- MLAT - Voortoets-dl. 2	- MLAT - Voortoets-dl. 2	- MLAT - Voortoets-dl. 2	
Sessie 3	- Herhaling studie-instructies - Leertaak 1 - Formatieve toets 1.1 - Feedback/totale herhaling - Formatieve toets 1.2 - Feedback/uitleg	- Herhaling studie-instructies - Leertaak 1 - Formatieve toets 1.1 - Feedback/partiële herhaling - Formatieve toets 1.2 - Feedback/uitleg	- Herhaling studie-instructies - Leertaak 1 - Formatieve toets 1.1 - Feedback/KCR - Formatieve toets 1.2 - Feedback/uitleg	- Herhaling studie-instructies - Leertaak 1 - Formatieve toets 1.1	 x x x x x
Sessie 4	- Leertaak 2 - Formatieve toets 2.1 - Feedback/totale herhaling - Formatieve toets 2.2 - Feedback/uitleg	- Leertaak 2 - Formatieve toets 2.1 - Feedback/partiële herhaling - Formatieve toets 2.2 - Feedback/uitleg	- Leertaak 2 - Formatieve toets 2.1 - Feedback/KCR - Formatieve toets 2.2 - Feedback/uitleg	- Leertaak 2 - Formatieve toets 2.1	 x x x x x
Sessie 5	- Tussentoets I - Leertaak 3 - Formatieve toets 3.1 - Feedback/totale herhaling - Formatieve toets 3.2 - Feedback/uitleg	- Tussentoets I - Leertaak 3 - Formatieve toets 3.1 - Feedback/partiële herhaling - Formatieve toets 3.2 - Feedback/uitleg	- Tussentoets I - Leertaak 3 - Formatieve toets 3.1 - Feedback/KCR - Formatieve toets 3.2 - Feedback/uitleg	- Tussentoets I - Leertaak 3 - Formatieve toets 3.1	 x x x x x
Sessie 6	- Herhaling studie-instructies - Leertaak 4 - Formatieve toets 4.1 - Feedback/totale herhaling - Formatieve toets 4.2 - Feedback/uitleg	- Herhaling studie-instructies - Leertaak 4 - Formatieve toets 4.1 - Feedback/partiële herhaling - Formatieve toets 4.2 - Feedback/uitleg	- Herhaling studie-instructies - Leertaak 4 - Formatieve toets 4.1 - Feedback/KCR - Formatieve toets 4.2 - Feedback/uitleg	- Herhaling studie-instructies - Leertaak 4 - Formatieve toets 4.1	 x x x x x

Sessie 7	- Tussentoets II	- Tussentoets II	- Tussentoets II	- Tussentoets II	
	- Leertaak 5	- Leertaak 5	- Leertaak 5	- Leertaak 5	x
	- Formatieve toets 5.1	- Formatieve toets 5.1	- Formatieve toets 5.1	- Formatieve toets 5.1	x
	- Feedback/totale herhaling	- Feedback/partiële herhaling	- Feedback/KCR		x
	- Formatieve toets 5.2	- Formatieve toets 5.2	- Formatieve toets 5.2		x
	- Feedback/uitleg	- Feedback/uitleg	- Feedback/uitleg		x
Sessie 8	- Leertaak 6	- Leertaak 6	- Leertaak 6	- Leertaak 6	x
	- Formatieve toets 6.1	- Formatieve toets 6.1	- Formatieve toets 6.1	- Formatieve toets 6.1	x
	- Feedback/totale herhaling	- Feedback/partiële herhaling	- Feedback/KCR		x
	- Formatieve toets 6.2	- Formatieve toets 6.2	- Formatieve toets 6.2		x
	- Feedback/uitleg	- Feedback/uitleg	- Feedback/uitleg		x
Sessie 9	- Leertaak 7	- Leertaak 7	- Leertaak 7	- Leertaak 7	x
	- Formatieve toets 7.1	- Formatieve toets 7.1	- Formatieve toets 7.1	- Formatieve toets 7.1	x
	- Feedback/totale herhaling	- Feedback/partiële herhaling	- Feedback/KCR		x
	- Formatieve toets 7.2	- Formatieve toets 7.2	- Formatieve toets 7.2		x
	- Feedback/uitleg	- Feedback/uitleg	- Feedback/uitleg		x
Sessie 10	- Attitudeschaal	- Attitudeschaal	- Attitudeschaal	- Attitudeschaal	
	- Natoets	- Natoets	- Natoets	- Natoets	
Sessie 11	- Retentietoets	- Retentietoets	- Retentietoets	- Retentietoets	

sessie 11 3 weken na sessie 10. In totaal nam de cursus dus plm. 9 weken in beslag. De proefpersonen waren vrij in de keuze van het uur waarop zij aan een leertaak wilden beginnen mits een en ander maar plaatsvond tussen 08.30 en 18.00 uur. Het onderzoek vond plaats in de onderzoeksruijnte (feedback-classroom) van de Vakgroep Onderwijspsychologie aan de Katholieke Hogeschool te Tilburg. Steeds waren één tot drie proefleiders aanwezig. In de eerste sessie werd de proefpersonen duidelijk gemaakt dat de cursus waaraan zij deelnamen weliswaar onderdeel uitmaakte van een experiment, maar dat zij toch in staat waren om door het serieus volgen van deze cursus hun kennis van de Franse taal weer op te frissen en/of uit te breiden. Aangegeven werd dat het gebruik van toetsen een essentieel element vormde van de cursus. In de introductie werd er voorts de nadruk op gelegd dat noch aan het bestuderen van de leertaken, noch aan het beantwoorden van de toetsvragen een tijdslimiet gekoppeld was. Gesteld werd dat de leertaken grondig bestudeerd moesten worden op dezelfde wijze als de proefpersonen gewoon waren te studeren. Pas wanneer zij serieus meenden de leerstof in voldoende mate te beheersen, mochten zij hun lesmateriaal inleveren en ontvingen zij een (formatieve) toets. Door aan de proefpersonen zelf over te laten te beoordelen of zij de leerstof al dan niet voldoende beheersten en 'aan de toets toe waren', werd een zo ecologisch valide mogelijke situatie geschapen. Tevens werd het interpretatieprobleem betreffende het stopcriterium (zie par. 7.5.2.1.) op deze wijze geminimaliseerd. Tijdens de cursus werden deze aanwijzingen nog eens herhaald in de sessies 3 en 6. Bij de proefpersonen werd er voorts op aangedrongen om gedurende de looptijd van de cursus geen extra studie aan het Frans te wijden, daar dit de resultaten van het experiment (ongunstig) zou kunnen beïnvloeden.

Na de eerste bijeenkomst werden de proefpersonen aselekt over de vier condities verdeeld, waarbij gestratificeerd werd naar geslacht. Het was de proefpersonen niet bekend aan welke conditie zij toebedeeld waren, noch in welke zin de condities zich onderling van elkaar onderscheidden.

In de sessies 5 en 7 werd een toets afgenomen die de doelstellingen van de voorafgaande twee leertaken moest meten (Tussentoets I resp. Tussentoets II). Na inlevering van deze toets bij de proefleider werd vervolgens weer een nieuwe leertaak aangeboden, conform de werkwijze in de eerder doorlopen sessies.

Sessie 10 werd geheel besteed aan het opnieuw invullen van de attitudeschaal en het maken van de natoets. Sessie 11 tenslotte werd gevuld met het maken van de retentietoets,

waarna het experiment voor de proefpersonen beëindigd was.

7.7. Design

Het design was een variant van het zgn. pretest-posttest-controle-groep-design (Campbell & Stanley, 1963), waarbij de proefpersonen na stratificatie naar de factor geslacht aselekt toegewezen werden aan de vier condities.

7.8. Data-verwerking

Ondanks dat de proefpersonen aselekt aan de condities zijn toebedeeld, is het toch zinvol om een voortoets af te nemen. Deze zinvolheid is niet alleen gelegen in het feit dat de voortoets in ons onderzoek een rol speelt met betrekking tot het cognitief begingedrag, maar wordt ook ontleend aan het gegeven dat de voortoets het mogelijk maakt na te gaan of er - ondanks de aselechte toewijzing - wellicht toch verschillen tussen de groepen bestaan. Voorts kan worden nagegaan in hoeverre de proefpersonen überhaupt iets geleerd hebben. Bovendien verhoogt afname van een voortoets het onderscheidingsvermogen van de statistische toets gebruikt voor de significantietoetsing van een mogelijk conditie-effect. In het onderhavige onderzoek is het bovendien ook inhoudelijk zinvol een voortoets af te nemen, daar de proefpersonen naar alle waarschijnlijkheid al een redelijke voorkennis hebben van het leerobject der cursus. Een en ander betekent dat de voortoets ook echt iets kan meten. De door de proefpersonen behaalde score is niet zonder meer een toevalscore.

Het opnemen van een voortoets in het onderzoeksdesign roept de vraag op naar de verwerking van de voortoetsgegevens. In de literatuur zijn in de loop der jaren vele adviezen gegeven aangaande deze verwerking. Simons (1981, pp. 108-112) geeft een overzicht van verschillende alternatieven voor de analyse van gegevens uit het voortoets-natoets design.

In het kader van dit onderzoek komen een viertal methoden in aanmerking. In het onderstaande worden deze kort besproken en wordt de keuze voor een dezer methoden beargumenteerd.

1. *Covariantie-analyse van de natoetsscores met de voortoets als covariant.*

Door middel van covariantie-analyse wordt op statistische wijze getracht de groepen of condities 'gelijk' te maken met betrekking tot hun voortoetsscores. Er wordt met andere woorden rekening gehouden met initiële verschillen in voor-

kennis, die een mogelijke verklaring zouden kunnen vormen van verschillen op de natoets. Cronbach & Furby (1970) merken op dat de voortoets als covariabele de nauwkeurigheid en het onderscheidingsvermogen van de statistische toetsing vergroot.

Elashoff (1969) en Kerlinger & Pedhazur (1973) stellen dat alleen dan covariantie-analyse gedaan mag worden, als de proefpersonen aselekt aan de verschillende condities zijn toegewezen en bovendien de regressie-coëfficiënten in de verschillende condities niet significant van elkaar verschillen (assumptie van de homogeniteit der regressie-coëfficiënten). Linn & Slinde (1977) tenslotte achten opname van de voortoets als covariabele alleen zinvol, als de correlatie tussen voor- en natoets minimaal .60 bedraagt.

2. *Regressie-analyse van de natoetsscores met de voortoets als predictor.*

(Multiple) regressie-analyse doet in feite hetzelfde als covariantie-analyse, terwijl het bovendien mogelijk is om ook nominale variabelen als predictor in de regressievergelijking op te nemen. In dat geval worden de nominale variabelen in de vorm van zgn. 'dummy-variabelen' als onafhankelijke variabelen in de regressievergelijking opgenomen.

3. *Variantie- of regressie-analyse op de natoetsscores.*

Bij aselechte toewijzing van de proefpersonen aan de condities raden verschillende auteurs aan (o.a. Cronbach & Furby, 1970; Linn & Slinde, 1977) om de voortoetsscores buiten beschouwing te laten en een variantie- of regressie-analyse op de natoetsscores uit te voeren. De voortoetscores vervullen dan slechts een controlefunctie. Simons (1981) merkt op dat deze methode een grotere interne en externe validiteit heeft, omdat er geen 'testing' en 'interaction effects of testing' kunnen optreden. Deze uitspraak is natuurlijk alleen waar als er ook inderdaad geen voortoets is afgenomen. Het wel of niet opnemen van de voortoets in de analyse doet niets ter zake. Bij het niet afnemen van de voortoets vervalt tevens de controlefunctie.

4. *Covariantie-analyse van de ruwe vershilscores met de voortoets als predictor.*

De methode is vergelijkbaar met methode 2). De vershilscores zijn echter beter interpreteerbaar dan voortoetscores.

Bovenstaande methoden kunnen nog verfijnd worden, door bij een onverhoopte lage betrouwbaarheid van de voortoets, niet te werken met de ruwe toetsscores, maar met geschatte ware scores van de proefpersonen op de voortoets.

Het uitvoeren van een variantie-analyse op de natoetsscores zonder daarbij gebruik te maken van de voortoets achten wij niet zinvol. In een eerdere analyse (Weeda, 1978) bleek namelijk dat voortoetsscores niet significant tussen groepen verschillen. Overeenkomstig de aanbeveling (methode 3) werd een variantie-analyse uitgevoerd op de natoetsscores. Een en ander leidde tot significante F-ratio's. Ter controle werd ook nog een covariantie-analyse uitgevoerd op de natoetsscores met de voortoets als covariant. De significante effecten bleken toen verdwenen te zijn. Het verdient derhalve geen aanbeveling om de voortoets slechts als controlemedium te gebruiken en niet te betrekken bij de analyse van de natoetsscores.

Gegeven de bovenstaande methoden gaat onze voorkeur uit naar methode 2): regressie-analyse op de natoetsscores met de voortoets als predictor. In het beslissingsschema van Simons (1981, p. 113) komt deze techniek ook als de meest geëigende naar voren, mits de correlatie tussen voor- en natoets $\geq .60$ en de voortoetsbetrouwbaarheid $\geq .80$ is.

De regressie-analyse is een zeer flexibele techniek en biedt ons tevens de mogelijkheid om de verschillende componenten van het model van Bloom als predictor voor leerresultaten op te nemen. Op eenvoudige wijze kan dan het percentage verklaarde variantie berekend worden, hetgeen ondermeer het doel was bij de validering van het model. In de paragrafen betreffende de resultaten zal verder ingegaan worden op de gehanteerde regressie-analyse.

7.9. Resultaten

7.9.1. Leerdoelgebonden toetsen

Voor de in het experiment gebruikte leerdoelgebonden toetsen, te weten de voortoets, tussentoets I, tussentoets II, natoets en retentietoets, werden betrouwbaarheidsindices berekend, gebaseerd op de totale onderzoeksgroep (dus niet uitgesplitst over de vier condities). De indices, tezamen met enige descriptieve gegevens zijn samengevat in tabel 7.4.

De betrouwbaarheidsindices zijn, met uitzondering van die voor TT-I, alle voldoende hoog. De KR-20 voor tussentoets I is, overigens niet zodanig laag, dat de scores alle betekenis verliezen.

In verband met vakanties e.d. namen uiteindelijk slecht 64 proefpersonen aan de afname van de retentietoets deel.

De in verband met de regressie-analyses van belang zijnde

intercorrelaties tussen de diverse toetsen zijn vermeld in tabel 7.5.

De correlaties van VT, TT-I en TT-II met de na- en retentietoets zijn zodanig dat deze toetsen in de uit te voeren regressie-analyses opgenomen kunnen worden (zie par. 7.8.).

Tabel 7.4. Betrouwbaarheidsindices en descriptieve gegevens leerdoelgebonden toetsen (N=84)

Toets	KR-20	M	SD
Voortoets (VT)	.95	65.79	21.89
Tussentoets I (TT-I)	.59	20.73	3.04
Tussentoets II (TT-II)	.73	16.76	3.12
Natoets (NT)	.95	106.37	18.22
Retentietoets (RT)*	.95	106.23	17.80

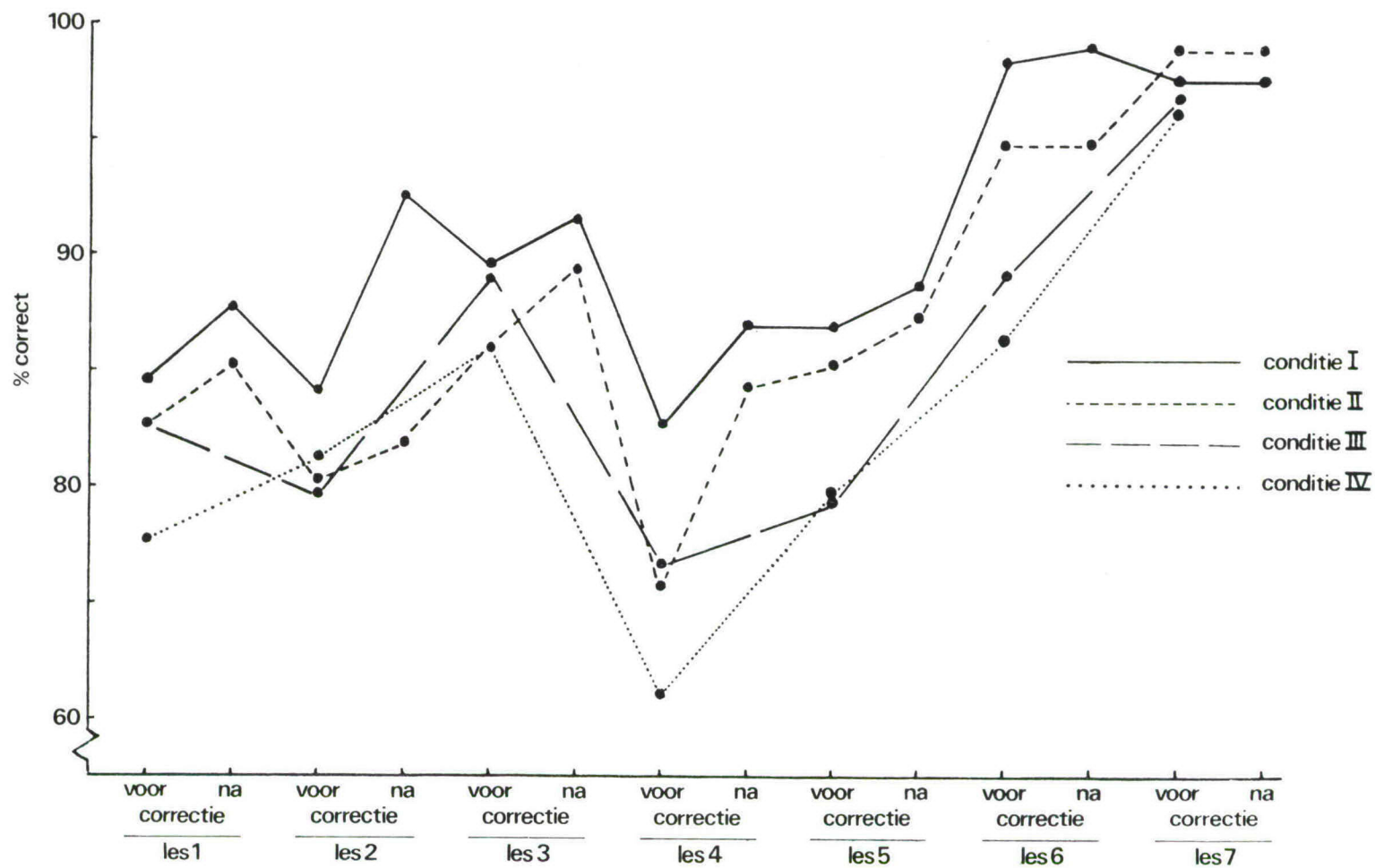
* Gebaseerd op N=64

Tabel 7.5. Correlatiematrix leerdoelgebonden toetsen

Toets	VT	TT-I	TT-II	NT	RT
Voortoets (VT)	1.00	.61	.58	.67	.66
Tussentoets I (TT-I)		1.00	.59	.61	.58
Tussentoets II (TT-II)			1.00	.76	.73
Natoets (NT)				1.00	.95
Retentietoets (RT)					1.00

7.9.2. Veranderingen in cognitieve leerprestaties over de leertaken

In het derde hoofdstuk van dit proefschrift werd in fig. 3.2. grafisch weergegeven hoe het prestatieverloop over vier leertaken in een mastery- en een non-mastery-conditie was in een experiment bij het leren van een tweede taal. Uit de grafiek bleek o.m. dat de leerprestaties op elke leertaak veranderbaar waren. Een soortgelijke grafiek is getekend voor de weergave van de veranderingen in leerprestaties over de 7 leertaken in ons experiment. In fig. 7.1. zijn leercurves getekend voor de vier condities. De grafieken voor de condities I en II kunnen beschouwd worden als vergelijkbaar met die van de mastery-conditie in fig. 3.2., terwijl die voor



Figuur 7.1. Veranderingen in de leerprestaties over diverse leertaken.

de condities III en IV overeenkomen met de non-mastery-conditie in fig. 3.2.

In fig. 7.1. is eerst voor elke conditie berekend wat het gemiddelde percentage goed beantwoorde items per formatieve toets was, waarna vervolgens voor de condities I en II een soortgelijk percentage berekend werd nadat een correctieve procedure had plaatsgevonden.

In grote lijnen vertoont fig. 7.1. hetzelfde beeld als in fig. 3.2. wordt geschetst. De hypothese van Bloom dat de leerprestaties op de diverse leertaken veranderbaar zijn wordt dus bevestigd. Duidelijk blijkt dat de leerprestaties na correctie voor de condities I en II gemiddeld hoger zijn dan de leerprestaties in de condities III en IV. Meestal, uitgezonderd bij de leertaken 2 en 3, is in de condities I en II ook de gemiddelde score op de formatieve toets vóór correctie hoger dan in de condities III en IV. Bloom veronderstelde dat voor de mastery-conditie telkens deze stijgende trend geconstateerd kon worden, terwijl voor de non-mastery-condities steeds een dalende trend zichtbaar zou worden (d.w.z. de gemiddelde score op formatieve toets 2 is lager dan die op formatieve toets 1, die op toets 3 is weer lager dan die op toets 2, etc.). Deze trend is niet aanwezig in fig. 7.1. Het niet-hiërarchische karakter van de leerstof is hier wellicht debet aan.

In het model van Bloom wordt ook gesteld dat de correlatie tussen leerprestaties na correctie en leerprestaties in de volgende leertaak vóór correctie afneemt in vergelijking met ongecorrigeerde leerprestaties in de voorafgaande leertaak (zie Hoofdstuk 3, par. 3.2.1.). In tabel 7.6. worden de door Bloom gerapporteerde mediane correlaties geplaatst naast de door ons gevonden mediane correlaties.

Tabel 7.6. Mediane correlaties tussen de leerprestaties op leertaak k en leertaak k+1.

	Volgens Bloom		dit onderzoek	
	mastery	non-mastery	mastery	non-mastery
vóór correctie	.47	.67	.22	.44
na correctie	.32	-	.17	-

Bij de berekening van de mediane correlatie-coëfficiënten in tabel 7.6 zijn de condities I en II samengevoegd en vormen aldus de mastery-conditie. Eveneens zijn de condities III en IV samengenomen, welke aldus als een non-mastery-conditie beschouwd kunnen worden. In de mastery-condities zowel bij

Tabel 7.7. Veranderingen in leerprestaties over leertaken voor mastery- en non-mastery proefpersonen.

	Leertaak 1				Leertaak 2				Leertaak 3			
	voor correctie		na correctie		voor correctie		na correctie		voor correctie		na correctie	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Mastery (ML)	25.11	3.58	26.00	2.85	19.77	1.99	21.02	1.98	19.34	2.57	19.91	1.76
Non-mastery (NM)	24.05	3.01			19.33	2.01			19.35	2.37		
ML/NM	104%				102%				100%			
s^2_{ML}/s^2_{NM}		141%				98%				118%		
	Leertaak 4				Leertaak 5				Leertaak 6			
	voor correctie		na correctie		voor correctie		na correctie		voor correctie		na correctie	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Mastery (ML)	17.57	2.85	18.86	2.24	67.18	6.70	68.61	5.56	18.88	1.24	18.96	1.15
Non-mastery (NM)	16.23	3.66			61.95	8.37			17.53	2.34		
ML/NM	108%				108%				108%			
s^2_{ML}/s^2_{NM}		61%				64%				28%		
	Leertaak 7											
	voor correctie		na correctie									
	M	SD	M	SD								
Mastery (ML)	11.77	.57	11.77	.57								
Non-mastery (NM)	11.63	.71										
ML/NM	101%											
s^2_{ML}/s^2_{NM}		64%										

Bloom als bij ons onderzoek, werd aan de leerlingen die daarvoor in aanmerking kwamen na afname van de formatieve toets extra tijd en instructie gegeven, terwijl dit in de non-mastery-conditie ontbrak.

Uit tabel 7.6. blijkt dat de op basis van het model van Bloom voorspelde trend in de afname van de correlatie vóór en na correctie aanwezig is, echter minder sterk dan bij de door Bloom gerapporteerde correlaties. Bij de interpretatie van de mediane correlaties dient men ook rekening te houden met de aantallen proefpersonen waarop deze gebaseerd zijn. In ons onderzoek zijn die resp. voor de mastery-conditie 44 en voor de non-mastery-conditie 40. De aantallen proefpersonen waarop Bloom zijn gegevens baseert zijn overigens van een gelijke grootte-orde. Een en ander betekent dat de correlatie-coëfficiënten nogal onderhevig kunnen zijn aan kansfluctaties. Het verschil tussen de gevonden correlaties vóór en na correctie zou hiernaan toegeschreven kunnen worden. De richting van de trend is echter geheel overeenkomstig het model. De range van de correlatie-coëfficiënten vóór correctie (van .10 tot .44) overlapt in ruimte mate die der coëfficiënten na correctie (van .10 tot .39). De range in de non-mastery-conditie varieert van .34 tot .49. Tenminste kan gesteld worden dat de gevonden relaties niet strijdig zijn met de predicties uit het model. Tevens blijkt wederom dat de leerprestaties na voltooiing van de leertaak veranderbaar zijn.

In tabel 7.7. worden de veranderingen in gemiddelden (M) en standaardafwijkingen (SD) vermeld voor de mastery- en non-mastery condities op de formatieve (her)toetsen in opeenvolgende leertaken. Voor elke leertaak worden de leerresultaten vermeld, behaald na de oorspronkelijke instructie (vóór feedback en correctiva) door beide groepen. Tevens zijn gegevens vermeld aangaande de leerresultaten van de mastery-groep na feedback en correctieve procedures. De gepresenteerde resultaten bevestigen het beeld zoals dat ook al in fig. 7.1. naar voren kwam. We zien een geringe stijging van de gemiddelde leerprestaties in de mastery-groep t.o.v. de andere groep. Voorts kan geconstateerd worden dat de relatieve variatie van de mastery-groep afneemt, terwijl die voor de non-mastery-groep juist toeneemt. Een en ander is in overeenstemming met de door Bloom (1976, p. 182-184) gerapporteerde empirische bevindingen.

7.9.3. Het cognitief begingedrag (CBG)

In tabel 7.8. zijn de gemiddelden en standaardafwijkingen van de scores op de voortoets en de aptitude-test (MLAT) vermeld.

Tabel 7.8. Gemiddelden en standaardafwijkingen m.b.t. het cognitief begingedrag

Conditie	Geslacht	Voortoets		MLAT	
		\bar{X}	SD	\bar{X}	SD
I	mn1. (16)	69.00	21.97	34.50	9.82
	vrl. (8)	76.88	20.63	34.88	7.52
II	mn1. (12)	72.17	22.52	31.65	7.60
	vrl. (8)	61.00	20.31	32.42	9.17
III	mn1. (13)	62.92	18.73	32.00	7.44
	vrl. (7)	72.57	9.43	31.62	6.42
IV	mn1. (12)	54.25	26.31	31.10	8.59
	vrl. (8)	65.25	23.38	29.17	11.21

Op de leerdoelgebonden voortoets behaalden de proefpersonen ongeveer een 48%-score. Er bleven derhalve nog genoeg mogelijkheden over om tijdens de cursus enige leerwinst te behalen. De geconstateerde verschillen tussen de vier condities bleken statistisch niet significant te zijn, zoals uit een tweeweg variantie-analyse bleek. (F-conditie (3;83) = 1.42, n.s.; F-geslacht (1;83) = .81, n.s. en F-interactie (3;83) = 1.14, n.s.). De geconstateerde verschillen op de MLAT bleken eveneens niet significant te zijn, noch m.b.t. de hoofdeffecten (F-conditie (3;83) = .77, n.s. en F-geslacht (1;83) = .30, n.s.), noch met betrekking tot het interactie-effect (F-interactie (3;83) = .51, n.s.).

De correlatie tussen voortoets en MLAT bedroeg .33 ($p < .01$). Ondanks dat er een geringe, te verwachten, positieve samenhang bestaat, blijken beide instrumenten toch elk verschillende componenten van het cognitief begingedrag te meten. Een en ander is in overeenstemming met het door Bloom gemaakte onderscheid.

7.9.3.1. Kanttekeningen bij de gehanteerde regressie-analyse techniek

Reeds eerder (par. 7.8.) is aangegeven dat voor de analyse van de onderzoeksresultaten gekozen is voor de techniek der multiple regressie-analyse met de leerdoelgebonden toetsen als afhankelijke variabelen en het cognitief begingedrag als onafhankelijke variabele (predictor).

Het is een bekend gegeven dat de volgorde waarin de predictoren in de regressie-analyse ingevoerd worden van invloed is op de unieke bijdragen van de variabelen aan de te verklaren variantie in de criteriumvariabele. Verandering van volgorde beïnvloedt de verdeling hiervan over de predictoren, zoals die wordt uitgedrukt in de gekwadrateerde semipartieële correlaties tussen de predictoren en de criteriumvariabele. De multiple correlatie zelf (het kwadraat hiervan geeft de proportie totaal verklaarde variantie aan) en de gewichten veranderen niet door verschillen in volgorde van predictoren.

In de door ons uitgevoerde multiple regressie-analyse is gekozen voor de zgn. *hierarchische methode*, waarbij de onderzoeker zelf specificeert in welke volgorde (groepen van) variabelen aan de regressievergelijking moeten worden toegevoegd. Door deze zelf gespecificeerde hiërarchie van cumulatieve invoering van predictoren houdt men in eigen hand op welke wijze de correlaties tussen de predictoren invloed uitoefenen op de regressievergelijkingen. Als de als tweede predictor ingevoerde variabele bijv. gecorreleerd is met de eerste predictor, dan kan de tweede predictor alleen nog maar aan de verklaring van de variantie in de criterium-variabele bijdragen met zijn unieke variantie (d.w.z. de niet-gemeenschappelijke variantie der twee predictoren). De volgorde van invoering kan dan wel het een en ander uitmaken. Het voor de statistische analyse gebruikte computerprogramma BMDP1R (Dixon & Brown, 1979) beschikt - zoals overigens de meeste regressieprogramma's - niet over de mogelijkheid de bijdrage van een predictor bij een dergelijke stapsgewijze analyse te toetsen. Aan de hand van de gegevens in de door het programma geconstrueerde overzichtstabel kan echter de bijdrage van elke predictor aan het percentage verklaarde variantie op significantie getoetst worden. Diverse auteurs geven hiervoor formules die equivalent aan elkaar zijn (cf. Kerlinger & Pedhazur, 1973, p. 124; Cronbach & Snow, 1977, p. 67; Van Knippenberg & Siero, 1980, p. 31).

Wij zijn uitgegaan van de formule voor de F-ratio:

$$F = \frac{(R^2_{y.12\dots k_1} - R^2_{y.12\dots k_2}) / (k_1 - k_2)}{(1 - R^2_{y.12\dots k_1}) / (N - k_1 - 1)}$$

waarin $R^2_{y.12\dots k_1}$ = gekwadrateerde multiple correlatie van de regressie van het criterium Y op k_1 variabelen;

$R^2_{y.12\dots k_2}$ = gekwadrateerde multiple correlatie van de regressie van het criterium Y op k_2 variabelen ($k_1 > k_2$);

k = aantal predictoren,

N = aantal waarnemingen,

(Van Knippenberg & Siero, 1980, p. 31).

De bijdrage van elke individuele predictor wordt dus getoetst tegen de variantie die resteert na het uitpartialiseren van alle predictoren.

In alle uitgevoerde regressie-analyses is gewerkt met niet-getransformeerde predictor- en criteriumscores. De literatuur geeft geen eenduidige aanwijzingen om wel of niet te transformeren, bijv. middels de z-transformatie. Cronbach en Snow (1977) bijv. waarschuwen tegen transformatie van criteriumvariabelen, maar achten een z-transformatie van predictoren niet onwenselijk. Kerlinger & Pedhazur (1973) citeren auteurs (p. 24) die zowel vóór als tegen transformaties van variabelen zijn. Zelf menen zij dat beide methoden in aanmerking moeten kunnen komen.

In de regressie-analyses waarbij het effect werd nagegaan van de conditie-variabele (een nominale variabele) werd gebruikt gemaakt van de zgn. dummy-codering. Voor de vier condities werden derhalve drie dummy-variabelen geconstrueerd (de vierde dummy-variabele bestond uit een vector met uitsluitend nullen). Kerlinger & Pedhazur (1973, p. 150) bevelen dummy-codering aan indien men werkt met verscheidene experimentele groepen en één controlegroep. Gezien de gehanteerde methode van dummy-codering zijn de (geringe) verschillen tussen aantallen proefpersonen in de verschillende condities ook niet zo bezwaarlijk als bij andere coderingsvormen, zoals bijv. bij orthogonale codering.

7.9.3.2. Voorspellende waarde van het cognitief begin-gedrag

Conform het in de voorgaande (sub)paragraaf gestelde zijn multiple regressie-analyses uitgevoerd teneinde na te gaan welk percentage van de variantie in leerprestaties verklaard kan worden door het CBG. Als criteriumvariabelen fungeren de leerdoelgebonden toetsen: TT-I, TT-II, NT en RT. Als predictoren fungeren de VT, de MLAT en de interactie VT x

MLAT. Op twee manieren is de voorspellende waarde van de beide componenten van het cognitief begingedrag nagegaan. In de eerste berekening is eerst de voortoets als predictor in de regressie-analyse ingevoerd en vervolgens de aptitudetest. Bij de tweede berekening is de omgekeerde volgorde aangehouden. Vanuit theoretische overwegingen kunnen argumenten aangevoerd worden om de MLAT vooraf te laten gaan aan de VT. Immers 'aptitude' is een min of meer constant kenmerk van een persoon, dat niet zo gemakkelijk beïnvloed (en dus veranderd) kan worden door leerprocessen. Overigens zij opgemerkt dat de correlatie tussen MLAT en VT niet zodanig is dat er een grote mate van gemeenschappelijke variantie is. Het volgorde-effect zal derhalve niet zo groot zijn.

In tabel 7.9. wordt een overzicht gegeven van de percentages door het CBG verklaarde variantie en de bijbehorende F-ratio's. Het bovenste gedeelte van de tabel geeft de resultaten weer waarbij eerst de VT en vervolgens de MLAT als predictor in de regressie-analyse is ingevoerd. Het onderste gedeelte van de tabel vermeldt de resultaten van de omgekeerde volgorde van invoering. Uit de in tabel 7.9. gepresenteerde gegevens blijkt dat de bijdragen van alle predictoren (incl. de interactie) significant zijn, met uitzondering van de interactieterm bij tussentoets I. Ook kan geconstateerd worden dat de door de voortoets verklaarde percentages variantie in alle gevallen hoger zijn dan die verklaard door de aptitudetest. De voortoets verklaart gemiddeld 32,4% van de variantie in leerprestaties, terwijl de MLAT gemiddeld 14,4% verklaart. Met name de door de voortoets verklaarde percentages variantie op de na- en retentietoets zijn hoog. Een mogelijke oorzaak hiervan is het gegeven dat deze toetsen identiek zijn. Het cognitief begingedrag als geheel verklaart over alle criteriumvariabelen gemiddeld 50,3% van de variantie in leerprestaties, waarbij vooral de percentages op de NT en de RT erg hoog zijn.

In het model van Bloom werd gesteld dat het CBG de verschillen in leerprestaties op een summatieve toets voor plm. 50% kon verklaren. De in dit experiment gevonden percentages zijn dus volledig in overeenstemming met het model. De hypothese met betrekking tot het CBG kan derhalve geconfirmeerd worden.

In tabel 7.10. zijn de ongestandaardiseerde regressie-coëfficiënten opgenomen betreffende de regressie van het CBG op verschillende criteriumvariabelen.

Bij de interpretatie van de regressie-coëfficiënten in tabel 7.10. dient men er rekening mee te houden dat de variabelen MLAT en VT op verschillende schalen gemeten zijn. De

Tabel 7.9. Percentages verklaarde variantie en F-ratio's met betrekking tot het cognitief begingedrag.

predictor	df	TT-1		TT-2		NT		df	RT	
		% var.	F	% var.	F	% var.	F		% var.	F
totale model	3	45.0	21.83**	39.5	17.37**	56.2	34.23**	3	60.4	30.56**
VT	1	37.7	54.88**	33.1	43.75**	45.3	86.34**	1	43.5	66.01**
MLAT	1	7.3	10.53**	3.9	4.74**	7.1	13.01**	1	9.4	14.23**
VT x MLAT	1	0.1	.08	2.8	3.63*	3.8	6.97**	1	7.5	11.45**
Residu	80	55.0		60.5		43.8		60	59.6	

totale model	3	45.0	21.83**	39.5	17.37**	56.2	34.23**	3	60.4	30.56**
MLAT	1	22.3	32.50**	14.8	19.53**	24.2	44.19**	1	25.9	39.25**
VT	1	22.6	32.91**	21.9	28.96**	28.2	51.53**	1	27.0	40.99**
MLAT x VT	1	0.1	.08	2.8	3.63*	3.8	6.97**	1	7.5	11.45**
Residu	80	55.0		60.5		43.8		60	59.6	

Tabel 7.11a. Percentage verklaarde variantie door het CBG voor de mastery-conditie (volgorde predictoren: VT, MLAT)

predictor	df	TT-1		TT-2		NT		df	RT	
		% var.	F	% var.	F	% var.	F		% var.	F
totale model	3	33.4	6.68**	33.6	6.75**	46.6	11.62**	3	40.4	7.24**
VT	1	27.4	16.42**	21.4	12.87**	33.0	24.70**	1	26.8	14.41**
MLAT	1	6.0	3.61	11.8	7.13*	12.3	9.23**	1	12.3	6.61*
VT x MLAT	1	.0	.00	0.4	.24	1.3	.93	1	1.3	.69
Residu	40	66.6		66.4		53.4		32	59.6	

* p < .05
** p < .01

Tabel 7.11b. Percentages verklaarde variantie door het CBG voor de non-mastery-conditie (volgorde predictoren: VT, MLAT)

predictor	TT-1			TT-2			NT		RT		
	df	% var.	F	% var.	F		% var.	F	df	% var.	F
totale model	3	62.5	19.98**	44.5	9.61**	62.2	19.73**		3	71.7	20.26**
VT	1	52.2	50.09**	42.2	27.32**	55.6	52.90**		1	55.7	47.22**
MLAT	1	9.7	9.37**	.2	.11	2.9	2.80		1	5.4	4.58*
VT x MLAT	1	.5	0.47	2.2	1.40	3.7	3.48		1	10.6	8.97**
Residu	36	37.5		55.5		37.8			24	28.3	

Tabel 7.12a. Percentages verklaarde variantie door het CBG voor de mastery-conditie (volgorde predictoren: MLAT, VT)

predictor	TT-1			TT-2			NT		RT		
	df	% var.	F	% var.	F		% var.	F	df	% var.	F
totale model	3	33.4	6.68**	33.6	6.75**	46.6	11.62**		3	40.4	7.24**
MLAT	1	13.6	8.18**	20.2	12.16**	23.5	17.58**		1	17.6	9.47**
VT	1	19.7	11.85**	13.0	7.83**	21.8	16.35**		1	21.5	11.56**
VT x MLAT	1	.0	.00	.4	.24	1.3	.93		1	1.3	.69
Residu	40	66.6		66.4		53.4			32	59.6	

Tabel 7.12b. Percentages verklaarde variantie door het CBG voor de non-mastery-conditie (volgorde predictoren: MLAT, VT)

predictor	TT-1			TT-2			NT		RT		
	df	% var.	F	% var.	F		% var.	F	df	% var.	F
totale model	3	62.5	19.98**	44.5	9.61**	62.2	19.73**		3	71.7	20.26**
MLAT	1	37.4	35.87**	11.3	7.34**	24.7	23.55**		1	36.4	30.88**
VT	1	24.6	23.58**	31.0	20.09**	33.8	32.17**		1	24.7	20.93**
VT x MLAT	1	.5	.47	2.2	1.40	3.7	3.48		1	10.6	8.97**
Residu	36	37.5		55.5		37.8			24	28.3	

* p < .05

** p < .01

Tabel 7.10. Ongestandaardiseerde regressie-coëfficiënten voor het cognitief begingedrag

criterium- variabele	Constante	b voor MLAT	b voor VT	b voor VT x MLAT
TT-1	13.56	.08	.06	.00
TT-2	3.79	.26	.15	-.00
NT	14.36	1.91	1.05	-.02
RT	-5.21	2.56	1.33	-.03

regressie-coëfficiënten kunnen dan niet zonder meer vergeleken worden.

Eveneens is nagegaan of er verschil bestond in voorspellend vermogen van het cognitief begingedrag ten aanzien van leerprestaties tussen de mastery-conditie enerzijds en de non-mastery-conditie anderzijds. Volgens de geformuleerde hypothesen zou het voorspellend vermogen van het CBG in de mastery-conditie lager moeten zijn dan in de non-mastery-conditie. De tabellen 7.11a,b en 7.12a,b geven een overzicht van de percentages verklaarde variantie door het CBG voor de mastery- en non-mastery-condities. Het bovenste gedeelte (a) van elke tabel heeft betrekking op de mastery-conditie; het onderste gedeelte (b) van elke tabel verwijst naar de non-mastery-conditie. In tabel 7.11a,b is eerst variabele VT in de regressie-analyse ingevoerd en vervolgens de variabele MLAT. In tabel 7.12a,b is deze volgorde omgedraaid.

Duidelijk blijkt uit beide tabellen dat het voorspellend vermogen van het CBG in de mastery-conditie lager is dan in de non-mastery-conditie. Gemiddeld verklaart het CBG voor de mastery-leerlingen 38.5% van de variantie op de afhankelijke variabelen, terwijl ditzelfde percentage voor de non-mastery-leerlingen 60.2% bedraagt. De conclusie die hieruit getrokken kan worden is dat de vormgeving van het onderwijsleerproces de invloed van voorafgaand leren op latere leerprestaties reduceert. Een bemoedigend gegeven voor de ontwerpers van onderwijsarrangementen!

Uit de tabellen 7.11 en 7.12 blijkt ook dat de voortoets een grotere verklarende waarde heeft met betrekking tot de variantie in leerprestaties dan de aptitude-test. De interactie tussen VT en MLAT levert alleen in de non-mastery-conditie ten aanzien van de retentietoets een significante F-ratio op.

Tabel 7.13 tenslotte geeft weer een overzicht van de

ongestandaardiseerde regressie-coëfficiënten uitgesplitst over de twee condities.

Tabel 7.13 Overzicht van de ongestandaardiseerde regressie-coëfficiënten per conditie voor het cognitief begingedrag

Criterion-variabele	Conditie	a	b(MLAT)	b(VT)	b(interactie)
TT-1	Mastery	13.72	.08	.06	.00
	Non-mastery	14.36	.05	.04	.00
TT-2	Mastery	9.79	.15	.07	-.00
	Non-mastery	3.40	.20	.18	-.00
NT	Mastery	43.67	1.34	.66	-.01
	Non-mastery	10.33	1.75	1.15	-.02
RT	Mastery	40.55	1.34	.70	-.01
	Non-mastery	-31.17	3.36	1.82	-.04

Opvallend in tabel 7.13. is het negatieve intercept voor de non-mastery conditie op de retentietoets. Voorts blijkt ook uit deze tabel dat de interactieterm in de regressievergelijking weinig betekenis heeft.

Tot nu toe is bij de analyse van het door het CBG verklaarde percentage variantie uitgegaan van de scores der proefpersonen op de diverse afhankelijke variabelen. Geen rekening is echter gehouden met de eventuele verschillen die er bestaan tussen de proefpersonen in de tijd benodigd om bepaalde leerprestaties te leveren. In par. 7.1.2. hebben wij gewezen op de leertijd als mogelijke concurrerende verklaring voor verschillen in leerprestaties. Voor alle proefpersonen is derhalve de totale leertijd bepaald, dat wil zeggen de tijd besteed aan het bestuderen van de leertaak, het maken van de formatieve toets, het (eventueel) bestuderen van de correctieve aanwijzingen, het (eventueel) maken van de formatieve hertoets en het (eventueel) ondergaan van feedback en uitleg. In tabel 7.14. wordt een overzicht gegeven van de totaal tijd per conditie, alsmede van diverse scores en indices zoals gedefinieerd in par. 7.5.2.1.

Het bovenste gedeelte van tabel 7.14 heeft betrekking op de natoets-scores, terwijl het onderste gedeelte betrekking heeft op de retentietoetsscores. Daar de aantallen proefpersonen die aan de natoets en de retentietoets deelnamen verschilden, verschillen ook de gemiddelde totaaltijden.

Tabel 7.14. Tijd- en efficiëntie-scores van de mastery- en non-mastery condities

	Mastery	Non-mastery
Gemiddelde totaaltijd	322.523	321.525
Absolute score (NT)	111.773	101.450
Efficiëntie-score	.371	.346
Absolute index	1.102	
Efficiëntie index	1.072	
Gemiddelde totaaltijd	325.695	309.679
Absolute score (RT)	109.889	101.143
Efficiëntie-score	.357	.352
Absolute index	1.086	
Efficiëntie index	1.014	

Opmerkelijk is dat de gemiddelde totaaltijd in de mastery-conditie slechts heel weinig hoger is dan in de non-mastery-conditie. Dit zou er op kunnen wijzen dat de mastery-leerlingen een wat efficiëntere studie-methode ontwikkelen dan de non-mastery-leerlingen. Zoals te verwachten was, zijn de absolute index (= de ratio van de absolute score van de experimentele en die van de controle-groep) en de efficiëntie-index (= de ratio van de efficiëntie-score van de experimentele en die van de controle-groep) groter dan 1.00. De mastery-leerlingen presteren dus beter en efficiënter dan de non-mastery-leerlingen.

De efficiëntie-scores (leerprestatie/leertijd) zijn als criterium-variabelen opgenomen in een multiple regressie-analyse met als predictoren VT, MLAT en VT x MLAT.

Ook nu weer zijn twee regressie-analyses uitgevoerd in verband met de verschillende volgorde van invoering der predictoren.

De resultaten van deze analyses zijn vermeld in tabel 7.15. Het bovenste gedeelte van de tabel vermeldt de volgorde voortoets-apitudetest; het onderste gedeelte van de tabel de omgekeerde volgorde.

In vergelijking met de regressie-analyses op de natoets- en retentietoetsscores (zie tabel 7.11a,b/7.12a,b) is het percentage verklaarde variantie voor beide condities gemiddeld iets lager. De correctie voor tijd vermindert dus het percentage verklaarde variantie enigszins. Desondanks blijft het percentage-zoals-gevonden in overeenstemming met

Tabel 7.15. Percentages verklaarde variantie in efficiëntiescores door het CBG

Predictor	Mastery-conditie						Non-mastery-conditie					
	Efficiëntiescore (NT)			Efficiëntiescore (RT)			Efficiëntiescore (NT)			Efficiëntiescore (RT)		
	df	% var.	F	df	% var	F	df	% var	F	df	% var	F
totale model	3	39.3	8.61**	3	41.1	7.45**	3	67.8	25.25**	3	62.6	13.36**
VT	1	37.6	25.22**	1	39.7	23.17**	1	39.1	42.33**	1	44.7	26.33**
MLAT	1	1.6	1.11	1	.2	.17	1	27.5	29.83**	1	16.4	9.67**
VT x MLAT	1	.0	.00	1	1.0	.67	1	1.2	1.33	1	1.5	.83
Residu	40	60.7		32	58.9		36	22.2		24	37.4	
totale model	3	39.3	8.61**	3	41.1	7.45**	3	67.8	25.25**	3	62.6	13.36**
MLAT	1	7.7	5.22*	1	1.9	1.00	1	56.9	61.67**	1	49.6	29.17**
VT	1	31.5	1.11**	1	38.2	22.33**	1	9.7	10.50**	1	10.4	6.83*
VT x MLAT	1	.0	.00	1	1.0	.67	1	1.2	1.33	1	1.5	.83
Residu	40	60.7		32	58.9		36	22.2		24	37.4	

* $p < .05$ ** $p < .01$

het percentage-zoals-voorspeld door het model opmerkelijk. Geheel in lijn met de voorspelling op grond van het model (zie par. 3.2.1.) is dat de verklarende waarde van de aptitudetest in de mastery-conditie sterk is afgenomen. Zowel als we dit percentage vergelijken met de gevonden percentages in de regressie-analyse op niet voor tijd gecorrigeerde scores (zie tabel 7.11^a en 7.12^a) voor de mastery-conditie, als in vergelijking met de percentages in de non-mastery conditie in tabel 7.15.

Het verband tussen aptitude en leerprestaties op summatieve toetsen neemt voor mastery-leerlingen dus duidelijk af. Voor de non-mastery-leerlingen lijkt de relatie aptitude-leerprestaties daarentegen groter te worden, gezien de percentages verklaarde variantie.

7.9.4. De affectieve leerlingkenmerken

In par. 7.3.2. van dit hoofdstuk is een overzicht gegeven van de vragen die aan de proefpersonen voorgelegd zijn ter bepaling van de affectieve aanvangskenmerken. De eerste serie peilt de mening van de proefpersonen aangaande voor het object van de cursus van belang zijnde waarden, terwijl de tweede serie de mening peilt aangaande de mate waarin zij menen dat deze cursus de realisering van die waarden mogelijk maakt. De attitudeschaal is zowel voorafgaand aan, als na afloop van het experiment afgenomen. De resultaten van beide afnames zijn vermeld in tabel 7.17. De range bij de eerste serie vragen loopt van +4 (zeer belangrijk) tot 0 (zeer onbelangrijk), die bij de tweede serie van +2 (zeer mee eens) tot -2 (zeer mee oneens).

Vergelijking van de scores voorafgaand aan het experiment met die na afloop van het experiment maakt duidelijk dat voor beide condities op het merendeel der vragen de tweede maal een lagere score behaald wordt.

Tabel 7.18. bevat de gemiddelde productscores op de attitude-schaal voorafgaand aan en na beëindiging van het experiment. Deze tabel biedt de mogelijkheid na te gaan of de cursus aan de verwachtingen van de proefpersonen heeft beantwoord. Daarbij moet onderscheid gemaakt worden tussen het verwachtingspatroon bij de aanvang van de cursus en de waardering van de cursus na afloop.

Uit tabel 7.18. blijkt dat de cursus niet aan de verwachtingen van de proefpersonen heeft voldaan.

In dit verband is het interessant na te gaan of deze teruggang in positieve houding ten opzichte van de cursus toegeschreven

Tabel 7.17. Gemiddelde vraag-scores attitudeschaal voor de mastery- (=42) en non-mastery-conditie (N=38).

vraag	Mastery-conditie				Non-mastery-conditie			
	Serie I Pre		Serie I Post		Serie I Pre		Serie I Post	
	M	SD	M	SD	M	SD	M	SD
1	3.0	.7	2.9	.7	2.6	1.0	2.4	1.0
2	3.0	1.1	3.0	1.0	2.7	.9	3.1	1.0
3	2.9	1.0	2.8	.9	2.3	1.1	2.2	1.1
4	2.5	1.1	2.4	1.1	2.4	1.2	2.4	1.1
5	2.0	1.1	2.2	1.1	1.9	1.0	2.3	1.2
6	3.7	.6	3.8	.4	3.7	.5	3.7	.5
7	3.4	.9	3.5	.8	3.6	.7	3.4	.8
8	2.7	1.1	2.5	1.1	2.6	1.1	2.5	1.0
9	3.5	.6	3.4	.7	3.2	.7	3.3	.7
10	2.3	1.0	2.2	1.1	2.1	1.2	2.0	1.2
11	2.6	1.0	3.1	.8	2.7	1.0	2.9	.9
12	3.3	.6	3.4	.6	3.1	.8	3.1	.9
13	2.9	1.0	2.9	.9	2.4	1.0	2.6	1.2
14	2.6	.9	2.7	.9	2.3	.9	2.1	1.2

Vraag	Mastery-conditie				Non-mastery-conditie			
	SerieII Pre		SerieII Post		SerieII Pre		SerieII Post	
	M	SD	M	SD	M	SD	M	SD
15	1.2	.8	.5	1.2	.8	.9	.4	1.1
16	.6	.8	.2	1.3	.2	1.1	.6	1.2
17	.6	.9	.8	.9	.4	.8	1.0	.8
18	.4	1.1	-.2	1.2	-.0	1.1	-.7	.9
19	.2	.9	-.7	1.0	-.0	1.2	-.9	1.1
20	1.0	.9	-.3	1.3	.4	1.2	-.4	1.2
21	.2	1.4	.4	1.1	.2	1.5	.3	1.4
22	-.3	1.1	-.7	1.0	-.4	1.2	-.8	1.0
23	.9	1.1	.2	1.0	.9	1.0	.3	1.2
24	.2	.8	.7	.8	.0	1.2	.8	1.1
25	.7	.9	.8	.9	.6	.8	.6	.8
26	1.5	.8	1.2	.7	1.1	1.1	.7	1.1
27	.6	1.1	.4	1.2	.7	1.2	.4	1.1
28	.5	1.2	.2	1.2	.1	1.2	-.2	1.3

Tabel 7.18. Gemiddelde produktscores attitudeschaal voor de mastery- (N=42) en non-mastery-conditie (N=38).

vragen- produkt	Mastery-conditie				Non-mastery-conditie			
	voor		na		voor		na	
	M	SD	M	SD	M	SD	M	SD
V01 * V15	3.8	2.8	1.7	3.5	2.2	2.8	1.6	3.0
V02 * V16	2.0	2.9	.9	4.1	1.0	3.2	2.2	4.0
V03 * V17	2.2	3.0	2.2	2.9	1.3	2.4	2.5	2.2
V04 * V18	1.0	3.0	-.3	3.1	.1	3.3	-1.8	2.6
V05 * V19	1.0	2.5	-.9	2.4	.1	2.6	-1.3	3.0
V06 * V20	3.7	3.5	-1.0	4.9	1.7	4.8	-1.5	4.5
V07 * V21	.8	5.0	1.6	4.1	.7	5.6	1.2	5.1
V08 * V22	-.2	2.9	-1.2	2.5	-.7	3.7	-1.9	3.0
V09 * V23	3.2	4.0	.6	3.7	2.9	3.1	1.2	4.3
V10 * V24	.8	2.2	1.6	2.3	.6	3.0	1.6	2.5
V11 * V25	2.3	3.3	2.4	3.1	1.8	2.6	1.7	2.6
V12 * V26	4.9	2.6	4.0	2.6	3.6	4.0	2.7	3.8
V13 * V27	2.3	3.4	1.1	3.6	2.6	3.7	1.6	3.0
V14 * V28	1.6	3.2	.7	3.0	.8	3.4	.3	3.3

moet worden aan het feit dat men minder waarde is gaan hechten aan de kennis van de Franse taal (serie I) dan wel aan het feit dat men minder vertrouwen stelt in de efficiëntie van de cursus (serie II). Het antwoord op deze vraag kan afgeleid worden uit de in tabel 7.17 vermelde gegevens.

Met uitzondering van de score op vraag 5 liggen de scores op de eerste serie vragen op de pre-toets (afname van de attitudeschaal voorafgaande aan het experiment) alle boven de 2.0, het neutrale punt op de waardenschaal. Ook op de post-toets (afname na het experiment) liggen deze scores, inclusief die voor vraag 5, boven de waarde 2.0. Een en ander betekent dat de proefpersonen tijdens het experiment niet minder waarde zijn gaan hechten aan de kennis van de Franse taal. Gezien het feit dat iemands waardepatroon niet zo gemakkelijk verandert, is dit op zich niet zo'n onverwacht resultaat. Zowel op de pre- als op de post-afname van de attitudeschaal bleken de proefpersonen grote waarde te hechten aan het in vraag 6 gestelde, het op vakantie in Frankrijk met de mensen kunnen praten. Ook vraag 7 betreffende de besteding van de vrije tijd behaalde een hoge score.

Als we kijken naar de verschuiving per vraag van de pre-toets ten opzichte van de posttoets dan valt op dat met name op

vraag 11 in de mastery-conditie een relatief grote wijziging optreedt, namelijk van 2.6 naar 3.1, terwijl deze toename in de non-mastery conditie beduidend geringer is. Deze vraag betreffende het regelmatig gebruik van studietoetsen heeft te maken met de instrumentele waarde van de cursus.

In de non-mastery-conditie daarentegen zien we dat de score-toename van pre- naar posttoets op vraag 2 relatief groot is. Ook deze vraag had te maken met de instrumentele waarde van de cursus, namelijk het je eigen gang kunnen gaan in een cursus. De proefpersonen in deze conditie kregen in vergelijking met die in de andere conditie relatief weinig toetsen en correctiva. Anderzijds blijkt uit de ongewijzigde score op deze vraag in pre- en posttoets voor de mastery-conditie dat de frequente toetsmomenten en correctieve procedures niet geleid hebben tot een wijziging in de reeds tamelijk hoge positieve waardering.

Over het geheel blijkt echter toch dat vormgeving en inhoud van de cursus kennelijk niet datgene bieden wat de proefpersonen ervan hadden verwacht. Gezien het feit dat voor beide condities een dergelijke trend waarneembaar is, kan ook niet gesteld worden dat de toevoeging van feedback en correctieve procedures verantwoordelijk is voor deze negatieve attitude-wijziging.

Conform de eerder uitgevoerde analyse is ook door middel van multiple regressie-analyse nagegaan welk percentage van de variantie in leerlingprestaties op de diverse criteriumvariabelen wordt verklaard door de variabele 'affectieve leerlingkenmerken'. In tabel 7.19. zijn de resultaten van de regressie-analyses vermeld zowel voor de groep als geheel, als voor de mastery- en non-mastery-conditie.

Tabel 7.19. Overzicht van de percentages verklaarde variantie door de variabele 'affectieve leerlingkenmerken'.

Criterium- variabele	Totaal			Mastery-conditie			Non-mastery-cond.		
	% var	F	df	% var	F	df	% var	F	df
TT-1	1.6	1.23	1/78	2.3	.94	1/40	.0	.01	1/36
TT-2	.1	.05	1/78	.1	.05	1/40	.7	.26	1/36
NT	.1	.07	1/78	.0	.01	1/40	.5	.19	1/36
RT	.2	.12	1/60	.3	.10	1/33	2.2	.55	1/25

De in tabel 7.19. vermelde percentages verklaarde variantie, alsmede de bijbehorende niet-significante F-ratio's, maken duidelijk dat in ons onderzoek de variabele 'affectieve

leerlingkenmerken' als predictor voor leerprestaties is te verwaarlozen. De gevonden percentages kunnen toegeschreven worden aan toevalsfluctuaties. Als zodanig steken zij nogal schril af bij het door Bloom gerapporteerde percentage van 25%. Noch in de totale proefpersonengroep, noch in de afzonderlijke condities is enige trend in de door het model van Bloom voorspelde richting te vinden. Evenmin zijn duidelijke discrepanties in percentages verklaarde variantie aanwezig tussen de beide condities.

Gezien de geringe percentages verklaarde variantie achten wij het niet zinvol om (ongestandaardiseerde) regressiecoëfficiënten e.d. te vermelden. Evenmin is een regressieanalyse uitgevoerd op de efficiëntiescores. Immers de dan te verwachten percentages zullen zeker niet hoger zijn dan de nu gevonden percentages verklaarde variantie.

7.9.5. *Kwaliteit van de instructie*

7.9.5.1. *Descriptieve gegevens*

De analyse van de mate waarin de variabele 'kwaliteit van de instructie' in staat was variantie in leerprestaties te verklaren, werd eveneens uitgevoerd met behulp van regressieanalyse. De descriptieve gegevens aangaande de criteriumvariabelen zijn vermeld in tabel 7.20.

Tabel 7.20. Descriptieve gegevens der criteriumvariabelen.

Criterium- variabele	Mastery-conditie			Non-mastery-conditie		
	M	SD	N	M	SD	N
TT-1	21.43	3.31	44	20.10	2.54	40
TT-2	17.66	2.25	44	15.85	3.56	40
NT	111.77	14.90	44	101.45	19.90	40
RT	109.89	14.07	36	101.14	21.14	28

7.9.5.2. *Hoofdeffecten*

De variabele 'kwaliteit van de instructie' is in dit onderzoek geoperationaliseerd als 'mastery' versus 'non-mastery'. Bovendien zijn er met betrekking tot de variabele feedback en

correctie een viertal onderscheidingen gemaakt. De vergelijking van de diverse feedbacktypen rapporteren we in een volgende paragraaf. Hier beperken we ons tot de verklarende waarde van de variabele kwaliteit instructie. In tabel 7.21. zijn deze percentages verklaarde variantie vermeld, waarbij onderscheid is gemaakt tussen mastery en non-mastery enerzijds en tussen de vier feedback-condities anderzijds.

Tabel 7.21. Overzicht van de percentages verklaarde variantie door de variabele 'kwaliteit van de instructie'.

Criterium- variabele	Mastery vs. Non-mastery Feedback-condities (4)			
	% var	F(1,82/1,62)	% var	F(3,80/3,59)
TT-1	4.9	4.22*	11.4	3.43*
TT-2	8.8	7.89**	11.3	3.38*
NT	8.2	7.33**	9.2	2.69*
RT	6.0	3.93	7.3	1.57

* $p < .05$

** $p < .01$

In de linkerhelft van tabel 7.21. zijn de percentages verklaarde variantie vermeld zoals die vergeleken moeten worden met de door Bloom gerapporteerde percentages. De rechterhelft van tabel 7.21. geeft de percentages na een meer gedifferentieerde analyse, waarbij de vier feedback-condities als dummy-variabelen in de multiple regressie-analyse zijn opgenomen (d.w.z. er werden drie dummy-variabelen geconstrueerd).

Duidelijk blijkt uit tabel 7.21. (linkerhelft) dat de door ons gevonden percentages beduidend lager zijn dan de op basis van het model van Bloom voorspelde percentages. Bloom voorspelde een percentage verklaarde variantie in de grootte-orde van 20-25%. Het door ons gevonden percentage (gemiddeld 7%) wijkt daar nogal aanzienlijk van af. Zelfs als men uitgaat van de nadere differentiatie in feedback-condities, dan nog wijkt het gevonden percentage verklaarde variantie in grote mate af van het voorspelde percentage.

De ongestandaardiseerde regressie-coëfficiënten voor de mastery- en non-mastery-condities zijn vermeld in tabel 7.22.

Tabel 7.22. Ongestandaardiseerde regressie-coëfficiënten betreffende de variabele 'kwaliteit van de instructie'

Criterium-variabele	intercept	b
TT-1	20.10	1.33
TT-2	15.85	1.81
NT	101.45	10.32
RT	101.14	8.75

Er is eveneens een multiple regressie-analyse uitgevoerd met kwaliteit van de instructie (mastery vs non-mastery) als predictor en de efficiëntiescores (zie tabel 7.14) als criteriumvariabelen. Te verwachten is dat het percentage verklaarde variantie zal afnemen, gezien het feit dat de mastery-conditie iets meer leertijd verbruikt heeft dan de non-mastery-conditie. Tabel 7.23 bevat de percentages verklaarde variantie.

Uit tabel 7.23 blijkt dat het percentage verklaarde variantie in vergelijking met de regressie-analyse op de absolute scores op na- en retentietoets wel zeer drastisch is verminderd. Gezien de niet significante F-ratio's kan het percentage de facto gelijk aan nul gesteld worden. Dit percentage kan niet vergeleken worden met door Bloom gerapporteerde gegevens, daar deze niet corrigeert voor verschillen in leertijd.

Tabel 7.23. Percentages verklaarde variantie op de efficiëntiescores met kwaliteit van de instructie als predictor.

Criterium-variabele	% var	F	df
Efficiëntiescore (NT)	1.0	.86	1/82
Efficiëntiescore (RT)	.1	.03	1/62

7.9.5.3. Variaties in feedback en correctie

In tabel 7.21. is in het rechter gedeelte aangegeven welke percentages variantie verklaard worden door de feedback-condities. Duidelijk blijkt dat het maken van onderscheid in

typen feedback en correctie leidt tot een hoger percentage verklaarde variantie dan in het geval van mastery versus non-mastery.

In par. 7.13. zijn enige vraagstellingen geformuleerd met betrekking tot feedback en correctieve procedures. In het navolgende zullen deze vragen successief worden beantwoord.

Allereerst is nagegaan of feedback en correctie leidt tot betere leerprestaties dan het achterwege laten van feedback en correctie. Voor de beantwoording van deze vraag worden de condities 1, 2 en 3 samengevoegd. Immers in deze condities is sprake van enigerlei vorm van feedback en correctieve procedures. Deze samengevoegde condities worden vervolgens vergeleken met de resterende conditie, waarin feedback en correctie ontbreken. Door middel van variantie-analyse (in dit geval equivalent aan toepassing van de t-toets) is het effect op de verschillende afhankelijke variabelen nagegaan. Tabel 7.24. bevat de resultaten van deze analyses.

Uitgaande van een significantieniveau van 5% blijkt tamelijk verrassend dat behalve bij Tussentoets 2 de groepen (conditie 1+2+3 vs. conditie 4) niet significant van elkaar verschillen. Alhoewel er onmiskenbaar wel een trend in de voorspelde richting waarneembaar is, blijkt feedback en correctie niet zonder meer tot significant betere leerprestaties te leiden. De vraag naar de effectiviteit van actieve versus passieve feedback en correctie wordt beantwoord door een variantie-analyse op de scores van de condities 1 en 2 versus conditie

Tabel 7.24. Overzicht variantie-analyses: geen feedback en correctie versus feedback en correctie.

Afhankelijke variabele	Bron	SS	df	MS	F	p <
TT-1	Fb/GFb	23.57	1	23.57	2.63	.109
	Residu	735.99	82	8.98		
TT-2	Fb/GFb	67.00	1	67.00	7.69	.007
	Residu	714.56	82	8.71		
NT	Fb/GFb	1027.74	1	1027.74	3.22	.076
	Residu	26196.55	82	319.47		
RT	Fb/GFb	1101.31	1	1101.31	3.58	.063
	Residu	19096.44	62	308.01		

Tabel 7.25. Overzicht variantie-analyses: actieve versus passieve feedback en correctie.

Afhankelijke variabele	Bron	SS	df	MS	F	p*
TT-1	Act./Pass.	16.09	1	16.09	1.68	ns
	Residu	593.35	62	9.57		
TT-2	Act./Pass.	18.47	1	18.47	2.49	ns
	Residu	460.89	62	7.43		
NT	Act./Pass.	1233.82		1233.82	3.98	.05
	Residu	19223.93		310.06		
RT	Act./Pass.	367.19	1	367.19	1.28	ns
	Residu	14109.55	49	287.95		

* Overschrijdingskansen > .10 zijn niet vermeld. In plaats daarvan wordt vermeld 'ns' (niet significant).

Tabel 7.26. Overzicht variantie-analyses: totale herhaling versus partiële herhaling.

Afhankelijke variabele	Bron	SS	df	MS	F	p
TT-1	Tot./Part.	46.97	1	46.97	4.65	.037
	Residu	423.83	42	10.09		
TT-2	Tot./Part.	2.46	1	2.46	.48	ns
	Residu	215.43	42	5.13		
NT	Tot./Pass.	233.35	1	233.35	1.05	ns
	Residu	9316.38	42	221.82		
RT	Tot./Pass.	.05	1	.05	.00	ns
	Residu	6925.50	34	203.69		

3. Immers de proefpersonen in de condities 1 en 2 moeten na de afname van de formatieve toets nog iets doen met de correctieve procedure. De correctie, totale of partiële herhaling, vraagt van de leerling een nieuw leerproces, een actief omgaan met de leerstof. In conditie 3 echter wordt van

de leerling slechts een passieve kennisname gevraagd. Hij krijgt te horen welke door hem gegeven antwoorden onjuist waren en wat de goede antwoorden zijn. Deze laatste situatie is in het reguliere onderwijs de meest voorkomende. In tabel 7.25 zijn de resultaten van de diverse variantie-analyses vermeld.

De vraag tenslotte of totale herhaling tot betere (of slechtere) leerresultaten leidt dan partiële herhaling, is beantwoord door de condities 1 en 2 met elkaar te vergelijken. De resultaten van de uitgevoerde variantie-analyses zijn vermeld in tabel 7.26.

Uit de in tabel 7.26 vermelde F-ratio's blijkt dat de ene correctieve procedure niet tot betere leerprestaties leidt dan de andere. Slechts op Tussentoets 1 blijkt totale herhaling tot hogere scores te leiden dan partiële herhaling. Op alle volgende toetsen tenslotte kan een dergelijk effect niet aangetoond worden.

Bloom (1976, p. 129) vermeldt de correlatie tussen de variabele 'kwaliteit van de instructie' en de diverse summatieve toetsen. Het is niet geheel duidelijk hoe die correlatie-coëfficiënten daar berekend worden.

Wij hebben de volgende procedure gevolgd. Bekend is dat zonder bezwaar de namen van categorieën van een nominale variabele veranderd kunnen worden; mits equivalentierelaties ongewijzigd blijven. In het bijzonder kunnen de namen vervangen worden door getallen. Het is dan mogelijk om bijvoorbeeld de produktmoment-correlatiecoëfficiënt te berekenen. De waarde daarvan is echter afhankelijk van de gekozen kwantificering van de categorieën. Om uit de oneindige verzameling kwantificeringen er één te kiezen, moet men een criterium formuleren. Het ligt in de rede als criterium te hanteren: de maximale waarde van de correlatie. In ons onderzoek zijn de gekozen kwantificeringen van de categorieën der nominale variabele gelijk aan (een positieve lineaire transformatie van) de binnen elke categorie behaalde gemiddelde score op de kwantitatieve variabelen (de summatieve toetsen).

Op de boven geschetste wijze zijn de produktmoment-correlatie-coëfficiënten berekend tussen de 4 condities enerzijds en de summatieve toetsen anderzijds. De resultaten zijn vermeld in tabel 7.27.

De gevonden correlaties stemmen overeen met de percentages verklaarde variantie zoals vermeld in tabel 7.21.

Tabel 7.27. Correlaties tussen 'kwaliteit van de instructie' (condities) en diverse leerdoelgebonden toetsen.

Leerdoelgebonden toets	pmc.
Voortoets	.22
Tussentoets 1	.34
Tussentoets 2	.34
Natoets	.30
Retentietoets	.27

Geconstateerd kan worden dat de correlaties tussen summatieve toetsen en kwaliteit van de instructie weliswaar hoger zijn dan die tussen voortoets en instructiekwaliteit, maar niet in die mate als door Bloom geprediceerd. Bloom vermeldt correlatie-coëfficiënten in de grootte-orde van .50 (o.c., p. 129 e.v.).

Tenslotte is ook nog nagegaan of de beantwoording van de vragen met betrekking tot feedback en correctie anders zou luiden als rekening gehouden wordt met de verschillen in leertijd tussen de diverse condities. De resultaten van deze analyses zijn vermeld in tabel 7.28.

Tabel 7.28. Overzicht variantie-analyses met betrekking tot feedback en correctie met de efficiëntie-scores als afhankelijke variabele.

Analyse	Afh.var.	Bron	SS	df	MS	F	p
Feedback vs. geen feedback	Effsc./NT	Conditie	.017	1	.017	1.098	ns
		Residu	1.250	82	.015		
	Effsc./RT	Conditie	.017	1	.017	1.515	ns
		Residu	.686	62	.011		
Actief vs. passief	Effsc./NT	Conditie	.002	1	.002	.159	ns
		Residu	.888	62	.014		
	Effsc./RT	Conditie	.005	1	.005	.417	ns
		Residu	.554	49	.011		
Totaal vs. partieel	Effsc./NT	Conditie	.004	1	.004	.310	ns
		Residu	.599	42	.014		
	Effsc./RT	Conditie	.001	1	.001	.066	ns
		Residu	.349	34	.010		

Uit de in tabel 7.28 vermelde resultaten van de diverse variantie-analyses blijkt, dat na correctie voor leertijd de spaarzame significante effecten die vóór correctie nog optraden geheel verdwenen zijn.

7.9.6. Het totale model

In de voorafgaande paragrafen is uitvoerig aandacht besteed aan het variantieverklarend vermogen van de diverse afzonderlijke variabelen uit het model van Bloom, t.w. het cognitief begingedrag, de affectieve leerlingkenmerken en de kwaliteit van de instructie. In deze paragraaf wordt nagegaan welk percentage van de variantie in leerlingprestaties op de diverse summatieve toetsen door het model als geheel, derhalve inclusief de verschillende interacties, kan worden verklaard. Gezien het feit dat de variabele affectieve leerlingkenmerken nagenoeg geen variantieverklarende waarde had, is er van afgezien om interacties van deze variabele en de resterende variabelen in het model op te nemen. Wederom is via multiple regressie-analyse het percentage verklaarde variantie bepaald. Gezien het feit dat de variabele kwaliteit van de instructie als predictor in de regressie-analyse is opgenomen, zijn geen afzonderlijke regressie-analyses uitgevoerd voor de mastery- en non-mastery-conditie. Voor de conditie-variabele is weer een dummy-variabele geconstrueerd, waarbij de mastery-conditie aangeduid is als een vector met enen, en de non-mastery-conditie als een vector met nullen.

De resultaten van de uitgevoerde regressie-analyses zijn vermeld in tabel 7.29.

Tabel 7.29. Overzicht van de door het totale model verklaarde percentages variantie op enige summatieve toetsen (predictoren: Apt., VT, VT * Apt., QI, QI * VT, QI * Apt., QI * VT * Apt., Attitude).

Afhankelijke variabele	df	% var	F	Niet opgenomen predictor
TT-1	7/73	46.2	8.96**	QI * VT * Apt
TT-2	7/73	46.3	8.98**	QI
NT	7/73	61.3	16.51**	QI
RT	7/54	65.3	14.49**	QI * Apt.

** $p < .01$

Het aantal vrijheidsgraden in tabel 7.29 is steeds 1 lager dan het totaal aantal in het model opgenomen predictoren, omdat in elke uitgevoerde regressie-analyse steeds één predictor niet in de regressievergelijking werd opgenomen. Welke dit is, is vermeld in de laatste kolom van tabel 7.29. De reden voor niet opname van een predictor is een te geringe tolerantie. Als een onafhankelijke variabele of predictor een hoge gekwadrateerde multiple correlatie heeft met de andere onafhankelijke variabelen, dan draagt deze predictor vrijwel niets bij aan de regressievergelijking. De tolerantie is nu gelijk aan 1 minus deze gekwadrateerde multiple correlatie. In het voor de data-analyse van dit onderzoek gebruikte BMDPIR-programma (Dixon & Brown, 1979, p.380 e.v.) wordt steeds voor elke in de regressievergelijking opgenomen variabele de tolerantie berekend. Indien deze tolerantie te laag is, of als door opname van de ene predictor de tolerantie van een reeds opgenomen predictor te laag wordt, dan blijft deze predictor buiten de regressievergelijking.

Vergelijking van tabel 7.29 met tabel 7.9, waarin de door het CBG verklaarde percentages variantie vermeld zijn, maakt duidelijk dat de opname van de predictoren QI en attitude, alsmede de diverse interacties niet een veel hoger percentage verklaarde variantie oplevert dan het CBG alleen. Het totale model verklaart op de summatieve toetsen meer variantie dan het CBG afzonderlijk en wel 1.2% op de TT-1, 6.8% op de TT-2, 5.1% op de NT en eveneens 5.1% op de RT. Gemiddeld derhalve iets minder dan 5%.

Het in dit onderzoek gevonden percentage verklaarde variantie op de natoets en de retentietoets varieert van ruim 60% tot ruim 65%. Op basis van het model van Bloom zou een percentage van plm. 90% verwacht mogen worden. Een discrepantie derhalve van 30 à 25%. Te verwachten valt dat bij een regressie-analyse met de efficiëntiescores als criteriumvariabelen de percentages verklaarde variantie nog sterker zullen afnemen. Tabel 7.30 bevat een overzicht van deze percentages.

Tabel 7.30. Overzicht van de door het totale model verklaarde percentages variantie ten aanzien van de efficiëntiescores (predictoren: zie tabel 7.29).

Afhankelijke variabele	df	% var	F	Niet opgenomen predictor
Eff.score/NT	7/73	56.0	13.26**	QI * Apt.
Eff.score/RT	7/54	52.4	8.48**	QI * VT * Apt.

**p < .01

Uit de in tabel 7.30 vermelde percentages blijkt inderdaad dat het percentage verklaarde variantie in leerlingprestaties op een summatieve toets wordt gereduceerd na correctie voor tijd. Deze gecorrigeerde percentages kunnen overigens niet vergeleken worden met door Bloom gepresenteerde gegevens daar deze niet corrigeert voor verschillen in leertijd. Over het geheel genomen blijkt echter dat het model in het door ons uitgevoerde onderzoek minder variantie verklaart dan door Bloom wordt gepostuleerd.

7.10. *Discussie*

Het is in zekere zin opmerkelijk, dat in de meeste experimenten met betrekking tot de strategie van beheersingsleren nauwelijks wordt nagegaan of de geoperationaliseerde en geïmplementeerde strategie ook daadwerkelijk zodanig functioneert als op grond van theoretische veronderstellingen mag worden verwacht. In het in dit hoofdstuk beschreven experiment is allereerst nagegaan of de door ons geconstrueerde strategie wel functioneert overeenkomstig de (theoretische) ideale strategie (zie bijv. Bloom, 1976, p. 176 e.v.). In hoeverre de door ons gebruikte leertaak voldoet aan de eisen die Bloom stelt, is een kwestie van (subjectieve) interpretatie. Op grond van de descriptie van de cursus in hoofdstuk 5, en op grond van de nadere bespreking van de leertaak in par. 7.4.1 menen we te kunnen stellen dat de oorspronkelijke leertaak voldoet aan de eisen die Bloom in deze stelt. Met betrekking tot de gehanteerde correctieve procedures zijn in feite twee typen onderscheiden, waarbij het ene type (totale herhaling) overeenkomt met de Kelleropvatting van een correctivum en het andere type met de Bloomopvatting. Block & Burns (1977) rekenen beide tot de correctieve procedures die in de strategie van beheersingsleren gebruikt kunnen worden.

Beschouwen we nu de empirische bevindingen, zoals weergegeven in par. 7.9.2. Uit de grafische weergave van de veranderingen in leerprestaties in de verschillende leertaken (zie fig. 7.12), alsmede uit de in tabel 7.7. vermelde data kan een aantal conclusies getrokken worden, t.w.

1. de gemiddelde leerprestaties van leerlingen op een leertaak kunnen verhoogd worden door toepassing van correctieve procedures;
2. de spreiding in de scores op de formatieve toetsen na correctie neemt af (dit is een logisch gevolg van het onder 1. gestelde, alsmede van het bereik van het meetinstrument);

3. de leerprestaties in de mastery-conditie vóór correctie zijn in de opeenvolgende leertaken beter dan in de niet-mastery-conditie;
4. de spreiding in de leerprestaties vóór correctie is in de mastery-conditie meestal lager dan in de niet mastery-conditie.

(Opgemerkt moet worden dat bovenstaande bevindingen niet statistisch getoetst zijn).

Ook deze empirische bevindingen zijn in overeenstemming met het door Bloom gepostuleerde.

Wel overeenkomstig de trend, maar niet in overeenstemming met de orde van grootte is de mediane correlatie tussen de leerprestaties na correctie en de leerprestaties in de volgende leertaak vóór correctie. De door ons geconstateerde afname is geringer dan door Bloom is voorspeld (zie tabel 7.6.). Ook de absolute grootte van de correlatie-coëfficiënten zowel voor de mastery- als voor de non-mastery-conditie is in ons onderzoek geringer dan Bloom voorspelt. Een tweetal oorzaken kan hieraan ten grondslag liggen. In de eerste plaats de relatief kleine steekproefomvang, waardoor de correlatie-coëfficiënten door toeval nogal kunnen fluctueren. In de tweede plaats kan de homogene samenstelling van de groepen genoemd worden. Immers hoe homogener de groepen, hoe geringer de correlaties zullen zijn.

Samenvattend kan gesteld worden dat de strategie van beheersingsleren zoals in dit onderzoek geoperationaliseerd valide is ten opzichte van het theoretische model.

Met betrekking tot de variantie-verklarende waarde van de diverse componenten van het model zijn de empirische resultaten gedeeltelijk niet in overeenstemming met de modelvoorspellingen.

Het door het CBG verklaarde percentage variantie is in overeenstemming met het model en de door ons geformuleerde hypothesen. Indien geen onderscheid gemaakt wordt tussen de mastery- en non-mastery-conditie wordt op de summatieve natoets ruim 56% van de variantie door het CBG verklaard. Het model voorspelde 50%. Wanneer de beide condities wel onderscheiden worden, dan kan geconstateerd worden dat de door het CBG verklaarde variantie in de mastery-conditie 46.6% bedraagt (conform de voorspelling: $< 50\%$) en in de non-mastery-conditie 62.2% (conform de voorspelling $> 50\%$). Dit verschil in percentage verklaarde variantie kan verklaard worden uit het verschil in kwaliteit van de instructie. Immers door een optimalisering van het onderwijsleerproces wordt de invloed van voorafgaand leren en de invloed van specifieke aanleg teruggedrongen. Hierbij is overigens nog geen rekening

gehouden met de mogelijke extra leertijd die afname van de formatieve toetsen en de toepassing van correctieve procedures vergt.

Indien wel rekening gehouden wordt met de verschillen in leertijd kan geconstateerd worden dat het percentage verklaarde variantie voor *beide* condities enigszins afneemt (althans: als we het gemiddelde percentage voor de natoets en de retentietoets nemen). Het verschil tussen beide condities wordt echter groter dan zonder tijdscorrectie. Dit is in tegenspraak met onze voorspelling, waarbij wij ervan uitgingen dat leertijd een verklaring vormde voor verschillen tussen de groepen op de summatieve toetsen.

Een verklaring voor het toenemend verschil in percentage door het CBG verklaarde variantie kan gelegen zijn in geringere invloed van de aptitude als verklarende variabele in de mastery-conditie. Deze geringere invloed wordt ook door het model van Bloom voorspeld. Overigens moet hierbij wel worden bedacht dat de invloed als verklarende variabele van de voortoets hierdoor wel toeneemt. Deze kennis, gemeten door de voortoets, is echter wel door middel van het onderwijs te beïnvloeden.

De affectieve leerlingkenmerken bleken in ons onderzoek geen variantie in leerprestaties op een summatieve toets te verklaren. Conform het model zou een percentage van 20 à 25 verwacht kunnen worden. De empirische bevindingen staan dus in contrast met het door het model voorspelde.

De vraag naar de oorzaak van dit verschil is niet zo gemakkelijk te beantwoorden. Wellicht is de oorzaak gelegen in het meetinstrument zelf. Immers gebruik is gemaakt van een *aangepaste* versie van een door Knibbeler (1976) ontwikkelde attitudeschaal. De aanpassingen waren echter zodanig dat niet verwacht mocht worden dat daardoor de betrouwbaarheid en validiteit van het instrument aangetast zouden worden. Slechts die vragen die specifiek betrekking hadden op de vormgeving van de cursus ontwikkeld door Knibbeler zijn gemodificeerd tot vragen met betrekking tot de door ons ontwikkelde cursus. Overigens baseert Bloom zijn voorspellingen met betrekking tot de variantie-verklarende waarde ook op gelijksoortige meetinstrumenten. In verreweg de meeste experimenten wordt gebruik gemaakt van specifiek voor dat experiment geconstrueerde attitudeschalen. Soms bestond het instrument slechts uit een zeer gering aantal vragen. Een minderheid van de onderzoekers heeft gebruik kunnen maken van gestandaardiseerde attitudeschalen (zie ook: Block & Burns, 1977, p.26-30).

De oorzaak kan ook gelegen zijn in de relatief korte duur van de cursus, waardoor de proefpersonen onvoldoende tijd kregen

om substantieel hun kennis van de Franse taal te verbeteren. Nu kon dit slechts ten aanzien van enige onderdelen der grammatica gebeuren. Mogelijkerwijze zou bij een meer langdurige cursus wel sprake kunnen zijn van een variantieverklarende waarde van de variabele 'affectieve leerlingkenmerken'.

Tenslotte kan een mogelijke oorzaak ook gelegen zijn in de homogeniteit van de groepen. Immers alle proefpersonen deden op vrijwillige basis mee en allen waren wat betreft hun kennis van de Franse taal (zie par. 7.2) vergelijkbaar.

Vooralsnog kan echter de voorspelling van Bloom dat de affectieve leerlingkenmerken 25% van de variantie in leerlingprestaties op een summatieve natoets kunnen verklaren, niet bevestigd worden.

Het percentage variantie dat door de variabele 'kwaliteit van de instructie' wordt verklaard, zou volgens het model 20-25% moeten bedragen. Wij vonden op de natoets een percentage van 8.8. Derhalve beduidend lager dan het voorspelde percentage.

Het is niet zonder meer duidelijk hoe dit verschil kan worden verklaard. Een mogelijke oorzaak zou kunnen zijn dat niet alle leertaken achteraf even geschikt waren om feedback en correctie toe te passen. Soms was de oorspronkelijke instructie kennelijk zo duidelijk (of, zo eenvoudig) dat vrijwel alle proefpersonen aan het gestelde beheersingscriterium voldeden. Er waren dan nauwelijks mogelijkheden om feedback en correctie toe te passen, zodat met betrekking tot deze leertaken nauwelijks verschil tussen de non-mastery- en de mastery-conditie gevonden kon worden. Vanuit een onderzoekstechnisch standpunt zou dit er wellicht voor pleiten om de oorspronkelijke instructie maar zo slecht mogelijk te maken, opdat de invloed van feedback en correctie zo groot mogelijk kan zijn. Vanuit een standpunt van ecologische validiteit en generaliseerbaarheid verdient een dergelijke aanpak echter niet de voorkeur. Voorts kan ook niet aangenomen worden dat de experimenten waarop Bloom zijn voorspellingen baseert, van dergelijke 'slechte' leertaken gebruik hebben gemaakt. Immers diverse leertaken behoorden tot het reguliere onderwijsprogramma.

Wanneer rekening gehouden wordt met de leertijd blijkt dat de variabele 'kwaliteit van de instructie' geen variantieverklarende waarde meer heeft. Gezien de toegepaste correctie ligt de verklaring hiervan voor de hand.

Bij de uitgevoerde regressie-analyse met betrekking tot alle variabelen uit het model blijkt het gevonden percentage van ruim 61% op de natoets beduidend achter bij het door Bloom geprediceerde percentage van 90. Rekening houdend met de

leertijd wordt een percentage van 56 gevonden.

Samenvattend kan gesteld worden dat binnen het door Bloom voorgestelde model de variabele CBG een zeer belangrijke plaats inneemt; ook, alhoewel in mindere mate, in de strategie van beheersingsleren. De instructie-variabele speelt ook een rol, maar als rekening gehouden wordt met de leertijd, dan blijkt de invloed van deze variabele aanmerkelijk geringer.

Het onderzoek naar de verschillende vormen van feedback heeft weinig resultaat opgeleverd. Enigszins verrassend bleek dat het toevoegen van feedback en correctie aan een onderwijsleerproces niet automatisch tot een verhoging van de leerprestaties leidt. Overigens zijn ook in eerdere onderzoeken dergelijke resultaten aangetoond (bijv. McDonald & Allen, 1962; Sullivan, Schutz & Baker, 1971; Wentling, 1973). Vaak echter bleek dat de oorspronkelijke leertaak reeds zodanig was dat de invloed van feedback daardoor geminimaliseerd werd. Ook was vaak sprake van een tamelijk hoge 'presearch availability' (Kulhavy, 1977), waardoor de tekst onvoldoende bestudeerd werd: de goede antwoorden waren te gemakkelijk te vinden.

Koppelen we nu een en ander terug naar de resultaten van ons onderzoek, dan valt op dat niet alle leertaken in gelijke mate voor de proefpersonen feedback en correctie noodzakelijk maakten. In de summatieve natoets (en retentietoets) zijn ook items opgenomen betreffende doelstellingen die kennelijk door de meeste proefpersonen zonder feedback en correctie bereikt konden worden. Het is niet uitgesloten dat mogelijke effecten van feedback en correctie daardoor worden versluierd. Voorts is het ook mogelijk dat leerlingen met een laag CBG meer profijt hebben van feedback en correctie dan leerlingen met een hoog CBG. In een nieuw experiment zou deze vraagstelling nader geëxploreerd dienen te worden (zie Hoofdstuk 8). Geen verschil kon worden aangetoond tussen totale en partiële herhaling, of tussen een Keller- en een Bloom-type correctieve procedure. Enigszins opmerkelijk is, dat dit verschil ook niet aangetoond kon worden als rekening gehouden werd met de benodigde leertijd. Wellicht was de totale tijdsduur van het experiment te kort om de invloed van de leertijd in dit geval duidelijk uit te laten komen.

Het verhaal wordt eentonig, maar ook ten aanzien van actieve versus passieve feedback kon geen significant verschil aangetoond worden. Ook niet als de invloed van de leertijd werd verdisconteerd.

Tenslotte moet nog worden gewezen op het feit dat de retentietoetsscores slechts heel weinig lager waren dan de natoetscores. Mogelijk speelt het gegeven dat zowel de voor-, de na- en de retentietoets gebruik maakten van dezelfde items hierbij een rol.

EXPERIMENT II: EEN TYPE 3 ONDERZOEK

8.1. Inleiding

In het in het voorgaande hoofdstuk beschreven experiment ging het erom na te gaan of het model van Bloom de aanspraken die het maakte met betrekking tot percentages verklaarde variantie in leerlingprestaties op een summatieve toets, kon waarmaken. In het in dit hoofdstuk beschreven experiment komen we nog in detail terug op de twee gebruikte varianten van correctieve procedures: totale versus partiële herhaling. De eerste vorm komt het meest overeen met de correctieve procedure zoals gepropageerd door Keller (zie: Hoofdstuk 2), de tweede zoals gepropageerd door Bloom (idem).

Het onderhavige experiment is een voorbeeld van het zgn. Type III-onderzoek (cf. Block & Burns, 1977; zie ook: Hoofdstuk 4), waarin nagegaan wordt welke effecten de beheersingslerenstrategie heeft op leerlingen met bijvoorbeeld verschillende cognitieve beginkenmerken. Het doel van dit experiment is het nagaan wat de effecten zijn van de beheersingslerenstrategie op leerlingen met verschillend cognitief begingedrag. Tevens wordt nagegaan of er wellicht een differentieel effect is van totale versus partiële herhaling op leerlingen met verschillend CBG, waarbij tevens rekening gehouden wordt met de moeilijkheidsgraad van de leerstof. Het experiment sluit aan bij een onderzoek van Burrows & Okey (1975) waarin werd nagegaan wat de effecten waren van mastery learning op studenten met verschillende aptitudes en uit verschillende 'grades'. De onderzoekers veronderstelden dat de invloed van de verschillen in aptitude op de leerprestaties zouden worden gereduceerd door de onderwijsmethode. Gewerkt werd met een variantie-analytisch design, waarbij de studenten met betrekking tot 'gradelevel' en 'aptitude' onderverdeeld werden. In het onderzoek werden vier condities onderscheiden. In conditie I bestudeerden proefpersonen individueel een leertaak. De leerkracht organiseerde een en ander en hield de vorderingen van de leerlingen bij. De proefpersonen in conditie II deden hetzelfde en kregen bovendien informatie betreffende de onderwijsdoelen. De proefpersonen in conditie III kregen daarenboven toetsitems met betrekking tot de doelstellingen ter oefening. De proefpersonen in conditie IV tenslotte kregen doelstellingen, oefenitems en bovendien een formatieve toets en correctie.

De vierde conditie, de beheersingslerenconditie, behaalde de beste resultaten. Zelfs proefpersonen met een lage aptitude in

deze conditie leverden hogere prestaties dan proefpersonen met een hoge aptitude in niet-mastery learning condities. Wel bleek dat conditie IV meer leertijd gebruikte dan de andere condities. De auteurs bevelen dan ook aan dat 'additional studies are needed to determine if the effects of mastery learning are maintained under more tightly controlled time conditions'. In het door ons uitgevoerde experiment is met deze aanbeveling rekening gehouden.

In een experiment van Swanson (1977) worden drie types correctieve procedures onderscheiden. Een correctivum die hij aanduidt als een Keller-procedure (= herhaling van dezelfde leerstof), een Bloom-procedure (= diverse vormen van remediële instructie) en een traditionele procedure (= geen remediële instructie). Vergelijking der drie methoden op een summatieve natoets maakt duidelijk dat er geen significant verschil geconstateerd kan worden tussen de Keller- en de Bloom-procedures, terwijl de Bloom-procedure significant beter is dan de traditionele procedure. Bij analyse van zogenaamde winstscores verdwijnt dit effect overigens weer. Een probleem bij dergelijk onderzoek blijft echter de vraag wat onder 'traditioneel' verstaan moet worden.

In een dissertatie-onderzoek van Tierney (geciteerd bij: Block, 1974) bleek de Bloom-procedure significant beter te zijn dan een Keller-correctieve procedure.

In diverse Type III-experimenten is ook aandacht besteed aan de relatie tussen de strategie van beheersingsleren en de moeilijkheidsgraad van de leerstof. Glassnapp, Poggio & Ory (1975) constateerden dat de mastery-learning strategie significant hogere resultaten opleverde met betrekking tot de hogere categorieën van de taxonomie van Bloom (cf. Bloom, 1956), dan met betrekking tot de lagere niveaus. Als interessant bijproduct van hun onderzoek kwam naar voren dat de proefpersonen uit de ML-conditie een grotere mate van 'test-anxiety' vertoonden bij de uiteindelijke summatieve toets in vergelijking met de proefpersonen uit de traditionele conditie.

Bij het tot nu toe verrichte onderzoek is weliswaar onderscheid gemaakt tussen typen correctieve procedures, maar is bij de toedeling van proefpersonen aan condities geen rekening gehouden met verschillen in aanleg en/of de moeilijkheidsgraad van de leerstof. Uit een ex post facto-analyse van de data uit ons eerste experiment, waarbij uitsluitend gewerkt werd met de proefpersonen met extreme scores, ontstond de indruk dat proefpersonen met een laag CBG meer baat hadden bij totale herhaling als correctieve procedure dan bij partiële herhaling. Voor proefpersonen met een hoog CBG kwam dit niet zo duidelijk naar voren.

Wij veronderstellen nu, dat leerlingen met een laag CBG meer fouten maken dan leerlingen met een hoog CBG. En bovendien, dat de oorzaak van die fouten een andere is. Te verwachten is dat fouten gemaakt door de relatief 'goede' leerlingen een tijdelijk en toevallig karakter hebben, soms veroorzaakt door onoplettendheid, soms door een de facto niet begrijpen van de leerstof. Fouten door onoplettendheid of iets dergelijks veroorzaakt, behoeven niet door correctieve procedures hersteld te worden. Die veroorzaakt door verkeerd begrip wel. In de regel zal dit dan afdoende kunnen geschieden door correctiva die op leerstofonderdelen betrekking hebben (partiële herhaling). Indien echter de leerstof als geheel tamelijk moeilijk is, dan kan totale herhaling tot betere resultaten leiden, omdat daar het verkeerd begrepen onderdeel in zijn context opnieuw geleerd kan worden. Fouten gemaakt door relatief 'slechte' leerlingen zullen minder vaak een toevallig karakter dragen. Veelal zal sprake zijn van verkeerd begrijpen van de leerstof. Voor deze leerlingen zou totale herhaling in het algemeen tot betere leerresultaten moeten leiden; zelfs ongeacht de moeilijkheidsgraad van de leerstof.

In het in dit hoofdstuk beschreven experiment wordt onderscheid gemaakt tussen totale en partiële herhaling (de 'treatment'), tussen hoog en laag cognitief begingedrag en tussen moeilijke en eenvoudige leertaken. In totaal worden zes condities onderscheiden: drie met een hoog en drie met een laag CBG. De eerste conditie krijgt steeds bij elke leertaak totale herhaling, de tweede partiële en de derde conditie krijgt totale herhaling bij moeilijke, en partiële herhaling bij eenvoudige leertaken.

8.2. Hypotheses

In aansluiting op het in de voorgaande paragraaf gestelde kunnen nu de onderstaande hypothesen geformuleerd worden. Als zodanig zijn zij een nadere uitwerking van de in hoofdstuk 4 geformuleerde algemene vraagstellingen.

1. Binnen de strategie van beheersingsleren behalen leerlingen met een hoog CBG hogere scores op een summatieve toets dan leerlingen met een laag CBG.
2. Totale herhaling als correctivum leidt voor leerlingen met een laag CBG tot hogere scores op een summatieve toets dan partiële herhaling, uitgezonderd bij eenvoudige leertaken. Voor leerlingen met een hoog CBG maakt de aard der correctieve procedure geen verschil.
3. Bij (relatief) moeilijke leertaken zal totale herhaling in

het algemeen tot betere leerprestaties leiden dan partiële herhaling.

4. Totale herhaling als correctivum leidt tot een toename van de leertijd in vergelijking met partiële herhaling.

Bij de bovenstaande hypothese 3 wordt in feite een ordinale interactie verondersteld, waarbij hoog-CBG-proefpersonen altijd een hoge summatieve score zullen behalen, ongeacht moeilijkheidsgraad der leerstof of aard der correctieve procedure. Leerlingen met een laag CBG behalen bij moeilijke leerstof alleen bij totale herhaling een hoge score.

Bij de formulering van de hypothesen is geen rekening gehouden met de leerefficiëntie. Te verwachten is dat totale herhaling meer tijd kost dan partiële herhaling. Een afweging van aard der correctieve procedure gekoppeld aan de moeilijkheidsgraad der leerstof zou wel eens de meest efficiënte procedure kunnen zijn, zowel voor leerlingen met een hoog als voor leerlingen met een laag CBG.

8.3. Proefpersonen

Aan dit onderzoek namen in totaal 83 proefpersonen deel, 37 vrouwelijke en 46 mannelijke proefpersonen. Reeds na de eerste sessie bleek dat 5 proefpersonen (4 mannelijke en 1 vrouwelijke) niet aan het gehele experiment konden deelnemen. Zij worden in de bespreking van het experiment dan ook verder buiten beschouwing gelaten.

Alle proefpersonen waren studierend aan de Katholieke Hogeschool Tilburg. Evenals bij het eerste experiment geschiedde ook nu weer de deelname op vrijwillige basis. Een deel der proefpersonen (plm. 40%) nam aan het experiment deel op basis van de binnen de Subfaculteit Psychologie gebruikelijke proefpersoonuren-regeling, een ander deel (60%) ontving een (geringe) financiële vergoeding. Ook nu weer waren het totaal aantal uren, alsmede het geldbedrag gefixeerd, derhalve onafhankelijk van het feit of de proefpersonen veel of weinig tijd aan het bestuderen van de lessen moesten besteden. De gemiddelde leeftijd van de proefpersonen was 22.24 jaar, de standaarddeviatie bedroeg 1.65. Alle proefpersonen hadden voorafgaand aan hun studie een VWO-opleiding gevolgd. In vergelijking met het eerste experiment was de nu gehanteerde steekproef homogener met betrekking tot diverse persoonskenmerken. Gezien de in het eerste experiment geconstateerde zeer geringe verschillen aangaande de extra studie van het Frans is in dit experiment geen informatie met betrekking tot dit aspect verzameld. Ook de, achteraf gebleken, grotere homogeniteit van deze steekproef kan beschouwd worden als een

ondersteuning van deze beslissing.

Voor de verdeling van de proefpersonen in de twee categorieën: hoog vs. laag cognitief begingedrag werd gebruik gemaakt van hun scores op de instrumenten gebruikt voor de meting van het cognitief begingedrag (zie voor een beschrijving van deze instrumenten, par. 8.4.). Gezien het percentage variantie dat door de voortoets (VT) in het eerste experiment verklaard werd, werd de hoofdingeling der proefpersonen hierop gebaseerd. De proefpersonen werden in hoog en laag verdeeld al naar gelang hun score onder of boven de mediane voortoetsscore lag. Op analoge wijze werd dit ook gedaan met betrekking tot hun scores op de Verbale aanlegtest (VAT) en de Modern Language Aptitude Test (MLAT).

In tabel 8.1. is vermeld hoe de uiteindelijke verdeling der proefpersonen eruit kwam te zien.

Tabel 8.1. Mediaansplitsing der proefpersonen met betrekking tot het cognitief begingedrag.

Voortoets Hoog: N=41	VAT-Hoog: N=28	MLAT-Hoog: N=24 (1) MLAT-Laat: N= 4 (2)
	VAT-Laat: N=13	MLAT-Hoog: N= 4 (3) MLAT-Laat: N= 9 (4)
Voortoets Laag: N=42	VAT-Hoog: N=13	MLAT-Hoog: N= 4 (5) MLAT-Laat: N= 9 (6)
	VAT-Laat: N=29	MLAT-Hoog: N=10 (7) MLAT-Laat: N=19 (8)

De proefpersonen in de categorieën (aangeduid door middel van een tussen haakjes geplaatst cijfer) (1), (2) en (3) werden beschouwd als behorend tot de categorie Hoog-CBG, die in de categorieën (8), (7) en (6) tot de categorie Laag-CBG. De proefpersonen in de resterende categorieën (4) en (5) werden aselekt toebedeeld aan de groepen Hoog- resp. Laag-CBG.

Het resultaat van een en ander was dat in de categorie Laag-CBG uiteindelijk 44 en in de categorie Hoog-CBG 39 proefpersonen terecht kwamen. Na stratificatie op de factor

geslacht, werden de proefpersonen vervolgens aselekt verdeeld over de zes condities (drie voor Hoog-CBG en drie voor Laag-CBG).

Van de vijf uitgevallen proefpersonen behoorden twee tot de categorie Hoog- en drie proefpersonen tot de categorie Laag-CBG.

8.4. Instrumenten voor de meting van het cognitief begin-gedrag

In het model van Bloom betreffende het leren op school spelen de affectieve leerlingkenmerken een belangrijke rol bij de verklaring van de variantie in leerlingprestaties op een summatieve toets. In het in het voorgaande hoofdstuk beschreven experiment bleek dat de verklarende waarde van de affectieve leerlingkenmerken (zoals door ons geoperationaliseerd en gemeten) nihil was.

Tevens is opgemerkt dat voor de meting van een attitude-verandering die verklarende waarde moet bezitten voor leerlingprestaties, wellicht een zich langer in de tijd uitstrekkend leerproces nodig is, dan in de door ons uitgevoerde experimenten mogelijk is.

Gegeven het zojuist gestelde, is besloten om in dit tweede experiment de meting van affectieve leerlingkenmerken achterwege te laten.

- *Leerdoelgebonden voortoets*

De in dit experiment gebruikte leerdoelgebonden voortoets was gelijk aan die uit het eerste experiment, met weglating van die items die betrekking hadden op doelstellingen die in de leertaken van het tweede experiment niet voorkwamen. Het totaal aantal items in de toets bedroeg 55.

- *Modern Language Aptitude Test (MLAT)*

Ook in het tweede experiment is gebruik gemaakt van twee subtests van de MLAT, t.w. subtest 4 'Words in sentences' en subtest 5 'Paired associates'. Beide subtests doen slechts een gering beroep op kennis van de Engelse taal.

- *Verbale Aanleg Testserie 1969 (VAT'69)*

Aangezien in het onderhavige experiment nagegaan wordt wat het effect is van bepaalde correctiva op proefpersonen met onderscheiden cognitief begingedrag werd het zinvol geacht om de engelstalige MLAT aan te vullen met een Nederlands-talige equivalent. Gebruik is gemaakt van een test uit de Verbale Aanleg Testserie 69. De VAT'69 vormt een onderdeel van een reeks tests voor de hogere intelligentieniveaus, welke aan de Vrije Universiteit te Amsterdam ontwikkeld worden. De verbale testserie bestaat uit drie tests, die -

zoals uit de intercorrelaties blijkt (cf. Drenth & Van Wieringen, 1969) - voldoende onafhankelijk zijn om aan iedere test afzonderlijk een stuk specifieke variantie binnen het gebied van de taalaanleg te kunnen toeschrijven. In dit onderzoek is gebruik gemaakt van de test 'Functies van woorden'. Dit is een Nederlandse bewerking van de sub test 'Words in sentences' uit de MLAT, die door ons eveneens is gebruikt.

In de onderhavige test gaat het om sensitiviteit voor grammaticale structuren. In een 40-tal opgaven, te voltooien in 25 minuten, wordt de onderzochte een tweetal zinnen gepresenteerd waarbij in de eerste een woord (of een woordcombinatie) is onderstreept en waarbij in de tweede zin een vijftal woorden (of woordcombinaties) is onderstreept en aangegeven met een letter A tot en met E. Het is de bedoeling dat de onderzochte uit de tweede zin dat woord (of die woordcombinatie) kiest dat dezelfde of de meest gelijkende grammaticaal-functionele betekenis heeft als dat onderstreepte woord uit een eerste zin (o.c. p.2).

De betrouwbaarheid van de test 'Functies van woorden' bedroeg .89 (KR-20) en .82 via de splitsingsmethode (o.c. p.12).

In overeenstemming met de in het eerste experiment gevolgde procedure wordt ook nu weer de combinatie van leerdoelgebonden voortoets, MLAT en VAT'69 test 'Functies van woorden' beschouwd als de operationalisatie van het cognitief begingedrag (CBG).

8.5. Vormgeving van het onderwijsleerproces

In het tweede experiment zijn vier van de zeven in het eerste experiment gebruikte lessen opnieuw gebruikt. Naar aanleiding van de opgedane ervaringen werden enige geringe wijzigingen met betrekking tot vormgeving en lay-out in de leertaken aangebracht. Voor het verloop van het experiment zijn deze echter niet essentieel. Gebruikt werden de lessen 1, 3, 4 en 5. De keuze van de lessen is gebaseerd op de in het eerste experiment geconstateerde moeilijkheidsgraad der lessen. Moeilijkheidsgraad is in dit verband geoperationaliseerd als het aantal proefpersonen dat na bestudering van de les voor de eerste maal in aanmerking komt voor remediële instructie; hetzij partieel, hetzij totaal. De oorspronkelijke lessen 1 en 4 bleken het moeilijkst, met respectievelijk 24 en 22 maal een benodigde remediële instructie. Het aantal voor de oorspronkelijke lessen 2, 6 en 7 benodigde correctieve

procedures was § 6. De volgorde van de lessen in het tweede experiment is gelijk aan die in het eerste experiment, waarbij de lessen 1, 3, 4 en 5 hernummerd worden tot resp. 1, 2, 3, en 4.

De gehanteerde terugkoppeling en correctieve procedures bestaan uit respectievelijk totale en partiële herhaling conform de in het eerste experiment gevolgde procedure. De normen voor de toewijzing van totale of partiële herhaling zijn dezelfde als in het eerste experiment.

8.6. Instrumenten voor de meting der cognitieve leerresultaten

In vergelijking met het eerste experiment is de in het tweede onderzoek gebruikte cursus ongeveer 40% korter. Het werd in dit verband dan ook niet zinvol geacht om tijdens de cursus ook nog leerdoelgebonden tussentoetsen af te nemen. Door middel van de opbouw der andere leerdoelgebonden toetsen (zie onder), alsmede door de strikt aangehouden item-doelstellingcorrespondentie is het voorts toch mogelijk om na te gaan of de doelstellingen der diverse leertaken bereikt worden.

- *Leerdoelgebonden natoets*

In afwijking van de in het eerste experiment gevolgde procedure bestond de leerdoelgebonden natoets in dit experiment voor de helft uit items die ook reeds in de voortoets waren opgenomen en voor de andere helft uit nieuwe vergelijkbare items. Elke doelstelling werd door minimaal twee items in de toets gerepresenteerd. De nieuwe items werden op de gebruikelijke manier door een vakdeskundige gescreend. Screening én itemdoelstellingcorrespondentie vormde wederom de basis voor de inhoudsvaliditeit.

De uit de voortoets overgenomen items werden in aselechte volgorde opgenomen in de natoets. Dit geldt overigens ook voor de nieuwe items. Het totaal aantal items in de leerdoelgebonden natoets bedroeg 110.

- *Leerdoelgebonden retentietoets*

De retentietoets was samengesteld uit de items van de natoets, waarbij alleen de volgorde der items op aselechte wijze gewijzigd was. Het aantal items bedroeg derhalve ook hier 110.

8.7. Meting van de leertijd

Op gelijke wijze als in het eerste experiment werd door de proefleider(s) voor alle proefpersonen door middel van een stopwatch de tijd in minuten (en seconden) bijgehouden die besteed werd door de proefpersonen aan

- het bestuderen van de leertaken,
- het uitvoeren van de (eventuele) correctieve procedures,
- het volgen van de eventueel benodigde mondelinge uitleg door de proefleider
- het maken van de leerdoelgebonden formatieve toetsen en formatieve hertoetsen.

Op deze wijze werd voor alle proefpersonen de tijd besteed aan het bestuderen van leertaken en correctiva en de tijd besteed aan het maken van de diverse formatieve toetsen, bepaald.

8.8. Procedure

In tabel 8.2. wordt een overzicht gegeven van de procedure zoals die in het tweede experiment tijdens de verschillende sessies voor de diverse condities is gevolgd. De eerste sessie vond plaats twee weken voorafgaand aan de leertaaksessies. Van deze laatste sessies vonden er per week steeds twee plaats, met minimaal twee dagen tussenruimte. De zesde sessie vond één week na sessie 5 plaats en de afname van de retentietoets vond plaats in sessie 7, drie weken na sessie 6. De totale cursus nam derhalve plm. zeven weken in beslag. Ook nu weer waren de proefpersonen vrij in de keuze van het uur waarop zij aan een leertaak wilden beginnen, mits een en ander maar plaatsvond tussn 08.30 en 18.00 uur. Het gehele experiment vond plaats in de feedback-classroom van de Vakgroep Onderwijspsychologie aan de Katholieke Hogeschool Tilburg. Steeds waren daarbij één tot drie proefleiders aanwezig.

In de eerste bijeenkomst werd de proefpersonen verteld dat in het experiment waaraan zij deelnamen enige lessen betreffende grammaticale constructies in de (geschreven) Franse taal aan de orde zouden komen. Duidelijk werd gemaakt dat noch aan het bestuderen van de leertaken, noch aan het maken van de formatieve (her)toetsen een tijdslimiet gekoppeld was. Ook in dit experiment werd dus de onbeperkte studietijdmethode (zie par. 7.5.2.1.) gehanteerd.

De proefpersonen werd erop gewezen dat de leertaken grondig bestudeerd moesten worden. Pas als zij meenden de leerstof in voldoende mate te beheersen, mochten zij hun lesmateriaal inleveren en ontvingen zij een (formatieve) toets. Met name deze studie-instructies werden in de tweede sessie nog eens

Tabel 8.2

Sessie	Conditie I/IV	Conditie II/V	Conditie III/VI	Tijdregistratie
Sessie 1	- Introductie - Voortoets - MLAT - VAT	- Introductie - Voortoets - MLAT - VAT	- Introductie - Voortoets - MLAT - VAT	
Sessie 2	- Herhaling studie-instructies - Leertaak 1 - Formatieve toets 1.1 - Feedback/totale herhaling - Formatieve toets 1.2 - Feedback/uitleg docent	- Herhaling studie-instructies - Leertaak 1 - Formatieve toets 1.1 - Feedback/partiële herhaling - Formatieve toets 1.2 - Feedback/uitleg docent	- Herhaling studie-instructies - Leertaak 1 - Formatieve toets 1.1 - Feedback/totale herhaling - Formatieve toets 1.2 - Feedback/uitleg docent	 x x x x x
Sessie 3	- Leertaak 2 - Formatieve toets 2.1 - Feedback/totale herhaling - Formatieve toets 2.2 - Feedback/uitleg docent	- Leertaak 2 - Formatieve toets 2.1 - Feedback/partiële herhaling - Formatieve toets 2.2 - Feedback/uitleg docent	- Leertaak 2 - Formatieve toets 2.1 - Feedback/partiële herhaling - Formatieve toets 2.2 - Feedback/uitleg docent	 x x x x x
Sessie 4	- Leertaak 3 - Formatieve toets 3.1 - Feedback/totale herhaling - Formatieve toets 3.2 - Feedback/uitleg docent	- Leertaak 3 - Formatieve toets 3.1 - Feedback/partiële herhaling - Formatieve toets 3.2 - Feedback/uitleg docent	- Leertaak 3 - Formatieve toets 3.1 - Feedback/totale herhaling - Formatieve toets 3.2 - Feedback/uitleg docent	 x x x x x
Sessie 5	- Leertaak 4 - Formatieve toets 4.1 - Feedback/totale herhaling - Formatieve toets 4.2 - Feedback/uitleg docent	- Leertaak 4 - Formatieve toets 4.1 - Feedback/partiële herhaling - Formatieve toets 4.2 - Feedback/uitleg docent	- Leertaak 4 - Formatieve toets 4.1 - Feedback/partiële herhaling - Formatieve toets 4.2 - Feedback/uitleg docent	 x x x x x
Sessie 6	- Natoets	- Natoets	- Natoets	
Sessie 7	- Retentietoets	- Retentietoets	- Retentietoets	

herhaald.

Na de eerste sessie werden de proefpersonen volgens de in paragraaf 8.3. beschreven methode aselekt over de zes condities verdeeld. Het was de proefpersonen niet bekend op welke wijze de condities van elkaar verschilden. De sessies 6 en 7 werden volledig besteed aan de afname van respectievelijk de natoets en de retentietoets.

8.9. Design en data-analyse

Evenals in het eerste experiment is de proefopzet van het onderhavige experiment een variant op het voortoets-natoets-controle-groep-design (Campbell & Stanley, 1963). Door de gehanteerde procedure van toewijzing van proefpersonen aan de diverse condities kan het design ook gezien worden als een variant op het extreme-groepen-design, waarbij dan de proefpersonen behorend tot de categorieën (4) en (5) in tabel 8.1. buiten de analyse gehouden moeten worden. Cronbach & Snow (1977, p. 512 e.v.) merken op dat het extreme-groepen-design bepaalde voordelen heeft in het zogenaamde aptitude-treatment-interaction-(ATI)onderzoek. Het is voor de data-analyse economischer om proefpersonen die zich aan de uiteinden van de aptitude-verdeling bevinden aan de diverse treatments toe te wijzen, omdat de proefpersonen rond het gemiddelde van die verdeling tamelijk weinig informatie verschaffen ten behoeve van de regressie-hellingen. Ook de power van de statistische analyse in het extreme-groepen-design is groter dan in andere designs (cf. Borich & Godbout, 1974). Anderzijds bevelen Cronbach & Snow (o.c. p.512) ook aan om in ATI-onderzoek minstens 100 proefpersonen per conditie te gebruiken. Bij het extreme-groepen-design kan met een geringer aantal proefpersonen volstaan worden. Vanwege tamelijk triviale redenen (onderzoeksbudget, 'beschikbare' proefpersonen, e.d.) konden wij niet meer proefpersonen gebruiken in het onderzoek dan gebeurd is. Het aantal gebruikte proefpersonen is derhalve slechts een zwakke spiegeling van het theoretisch wenselijke aantal. De beide auteurs bevelen ook aan dat ruwe natoets- (en retentie-)scores als afhankelijke variabelen gehanteerd moeten worden. Voorts mogen onder geen beding de scores binnen treatments gestandaardiseerd worden (p. 27). Het gebruik van voortoetsscores als mogelijke interactie-variabelen wordt aanbevolen. Als analyse techniek prefereren zij regressie-analyse boven variantie-analyse (p. 514-515), alhoewel de laatstgenoemde bij het extreme-groepen-design het meest logisch voor de hand ligt.

Met betrekking tot de rapportage van de onderzoeksresultaten

merken Cronbach & Snow op (p. 514-516) dat zoveel mogelijk descriptieve gegevens vermeld moeten worden en dat het bovendien ook zinvol is om niet significante interacties te publiceren, daar het anders onmogelijk wordt voor vakgenoten om ook erg zwakke trends, geringe effectgroottes en dergelijke te accumuleren. De vele door Cronbach & Snow gepresenteerde secundaire data-analyses, alsmede de vele resultaten van meta-analyse van onderzoeksgegevens (cf. Glass, 1976, 1977) tonen de zinvolheid van dit advies aan.

8.10. Resultaten

8.10.1. Descriptieve gegevens

Voor de in dit experiment gebruikte leerdoelgebonden toetsen werden indices voor betrouwbaarheid berekend gebaseerd op de totale onderzoeksgroep (niet uitgesplitst over de verschillende condities). De indices, tezamen met enige descriptieve gegevens zijn samengevat in tabel 8.3.

Tabel 8.3. Betrouwbaarheid en descriptieve gegevens leerdoelgebonden toetsen (N=78).

Toets	KR-20	M	SD
Voortoets (VT)	.88	24.85	9.04
Natoets (NT)*	.94	82.82	18.21
Retentietoets (RT)	.93	81.31	18.08

* Gebaseerd op N=77.

De interne consistentie-indices zijn voor alle toetsen voldoende hoog. Bij de interpretatie van gemiddelden en standaardafwijkingen in tabel 8.3. moet rekening gehouden worden met het feit dat zowel de na- als de retentietoets twee maal zoveel items bevatten als de voortoets. De gegevens van één proefpersoon ontbraken bij de natoets.

De intercorrelaties tussen de diverse toetsen zijn vermeld in tabel 8.4.

Uit de intercorrelaties tussen VT en respectievelijk NT en RT blijkt dat de toegevoegde nieuwe items geen nadelige invloed hebben op de intercorrelaties. De gevonden waarden zijn zelfs hoger dan die in het eerste experiment, waarbij voor-, na- en

Tabel 8.4. Correlatie-matrix leerdoelgebonden toetsen

Toets	VT	NT	RT
Voortoets (VT)	1.00	.76	.74
Natoets (NT)		1.00	.95
Retentietoets (RT)			1.00

retentietoets identiek waren. Ook de bij de retentietoets toegepaste randomisatie van de itemvolgorde ten opzichte van de natoets heeft in vergelijking met het eerste experiment geen invloed gehad op de hoogte van de gevonden inter-correlatie.

Aangezien in dit experiment meerdere aptitude-metingen verricht zijn, is het interessant na te gaan in hoeverre deze gecorreleerd zijn. Tabel 8.5. bevat de diverse inter-correlaties.

Tabel 8.5. Correlatiematrix aptitude-metingen.

Toets	VAT	MLAT/4	MLAT/5	MLAT/Tot.
VAT	1.00	.69	.21	.62
MLAT (Subtest 4)		1.00	.10	.77
MLAT (Subtest 5)			1.00	.72
MLAT (Totaal)				1.00

De gegevens in tabel 8.5. maken duidelijk dat de samenhang tussen de VAT en subtest 4 van de MLAT bevredigend hoog is. De VAT is immers de Nederlandstalige bewerking van deze subtest. De correlatie tussen beide MLAT-subtests is, zoals gewenst, laag. Beide meten derhalve verschillende aspecten van hetzelfde aanlegdomein.

In tabel 8.6. tenslotte wordt een overzicht gegeven van gemiddelden en standaarddeviaties op diverse variabelen uitgesplitst naar de onderscheiden condities.

Het bovenste gedeelte van tabel 8.6. bevat de condities met een hoog CBG, het onderste gedeelte de condities met een laag CBG. Door middel van variantie-analyse is nagegaan of de verschillen in VT-, VAT- en MLAT-scores binnen de hoog-CBG-condities en binnen de laag-CBG-condities significant waren. Geen der variantie-analyses leverde een significante F-ratio

op, zodat de initiële verschillen toegeschreven kunnen worden aan toevalsfluctuaties.

Tabel 8.6. Gemiddelden en standaarddeviaties in de zes condities

	Totale herhaling		Partiële herhaling		Totaal/Partieel	
Toets	Conditie I (N=12)		Conditie II (N=12)		Conditie III (N=13)	
	M	SD	M	SD	M	SD
VT	30.17	4.61	34.83	8.27	31.15	6.30
VAT	32.42	5.05	32.33	4.76	30.08	4.80
MLAT	41.33	9.06	37.33	5.63	38.46	6.62
NT	93.17	12.46	95.83	14.35	93.69	17.47
RT	90.67	12.93	94.00	14.02	91.00	18.16
	Conditie IV (N=13)		Conditie V (N=15)		Conditie VI (N=13)	
	M	SD	M	SD	M	SD
VT	18.92	5.19	19.53	4.39	16.46	5.71
VAT	25.62	3.50	26.60	5.25	23.62	6.21
MLAT	32.54	5.36	31.53	6.44	30.46	5.98
NT	75.00	12.46	76.50*	13.14	65.00	16.11
RT	74.77	11.63	74.60	12.24	65.54	20.02

* Gebaseerd op N=14.

8.10.2. Toetsing der hypothesen

De verwachting was dat leerlingen met een hoog cognitief begingedrag beter zouden presteren op na- en retentietoets dan leerlingen met een laag cognitief begingedrag, ongeacht de aard van de correctieve procedure (Hypothese 1). Voorts werd verondersteld dat voor leerlingen met een laag cognitief begingedrag totale herhaling als correctivum tot hogere leerprestaties op na- en retentietoets zou leiden dan partiële herhaling, terwijl voor leerlingen met een hoog CBG geen verschil geconstateerd zou worden (Hypothese 2). Ook werd verondersteld dat bij moeilijke leertaken totale herhaling beter zou zijn dan partiële, zeker voor wat betreft de leerlingen met een laag-CBG (Hypothese 3). Tenslotte werd verondersteld dat totale herhaling meer leertijd zou vergen dan partiële herhaling. (Hypothese 4).

In tabel 8.7. zijn de resultaten vermeld van variantie-analyses op respectievelijk na- en retentietoetsscores met als factoren cognitief begingedrag (twee niveaus: Hoog/Laag) en condities (drie niveaus: TTTT, PPPP, TPTP).

Tabel 8.7. Resultaten variantie-analyses op na- en retentietoetsscores

Afh.var.	Bron	SS	df	MS	F	p <
Natoets (N=77)	CBG	9344.75	1	9344.75	44.65	.01
	Conditie	633.08	2	316.54	1.51	ns
	Interactie	428.47	2	214.24	1.02	ns
	Residu	14859.60	71	209.29		
Retentie- toets (N=78)	CBG	7954.39	1	7954.39	34.74	.01
	Conditie	511.76	2	255.88	1.12	ns
	Interacties	299.52	2	149.76	.65	ns
	Residu	16487.81	72	229.00		

Uit tabel 8.7. blijkt dat hypothese 1 bevestigd wordt. Leerlingen met een hoog-CBG presteren op een summatieve toets beter dan leerlingen met een laag-CBG. Dit betekent dus dat de strategie van beheersingsleren er in dit geval niet in slaagt de invloed van voorafgaand leren (volledig) uit te schakelen. Gezien het in experiment 1 gevonden percentage door het CBG verklaarde variantie geen opmerkelijk resultaat.

Het hoofdeffect condities levert geen significante F-ratio op, noch op de na-, noch op de retentietoets. Dit resultaat is in overeenstemming met de bevindingen van experiment 1 waar eveneens geen significant verschil tussen totale en partiële herhaling geconstateerd kon worden. De interactie tussen cognitief begingedrag en condities bleek eveneens niet significant te zijn. Dit resultaat is in tegenspraak met het in hypothese 2 gestelde. Verwacht was dat tenminste sprake zou zijn van ordinale interactie. Deze hypothese moet dus verworpen worden.

De interactie hypothese is ook nog, overeenkomstig de aanbevelingen van Cronbach & Snow, getoetst met behulp van de multiple regressie-analyse. Hierbij is gebruik gemaakt van een door Ward & Jennings (1973, p.113-130) beschreven procedure, waarbij een 'restricted model' getest wordt tegen een 'full model'. In het volledige model wordt er van uitgegaan dat er interactie bestaat tussen CBG en condities, in het beperkte model wordt geen interactie verondersteld. Het volledige model

(Model I) luidt:

$$Y = a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + E_1 \quad (1)$$

waarbij Y = vector van natoetsscores
 $X_1 \dots 6$ = conditie dummy's ($X_1=1$ voor conditie 1 en 0 voor de overige condities, etc.)
 $a_1 \dots 6$ = schatter (sub)populatie gemiddelde conditie 1....6

Voor het beperkte model (Model II) geldt, dat er geen interactie is tussen CBG en conditie. Dit impliceert:

$$\begin{aligned} a_1 - a_4 &= c_1 \rightarrow a_1 = a_4 + c_1 \\ a_2 - a_5 &= c_1 \rightarrow a_2 = a_5 + c_1 \\ a_3 - a_6 &= c_1 \rightarrow a_3 = a_6 + c_1 \end{aligned}$$

Substitutie van (2) in (1) levert op:

$$\begin{aligned} Y &= a_4(X_1+X_4) + a_5(X_2+X_5) + a_6(X_3+X_6) + c_1(X_1+X_2+X_3) + E_2 \\ &= b_1X^{(1)} + b_2X^{(2)} + b_3X^{(3)} + c_1X^{(h)} + E_2. \end{aligned}$$

waarbij $X^{(1)}$ = scores op treatment i
 $X^{(h)}$ = scores voor leerlingen met hoog CBG.

De toetsing van Model II tegen Model I geschiedt met behulp van onderstaande F-ratio:

$$F = \frac{(SSE_2 - SSE_1) / (n_1 - n_2)}{SSE_1 / (N - n_1)},$$

waarbij SSE_1 = error-kwadratensom Model I
 SSE_2 = error-kwadratensom Model II
 n_1 = aantal lineair onafhankelijke vectoren Model I
 n_2 = aantal lineair onafhankelijke vectoren Model II
 N = aantal proefpersonen.

De bijbehorende vrijheidsgraden zijn respectievelijk $df_1 = n_1 - n_2$ en $df_2 = N - n_1$.

Uitvoering van beide multiple regressie-analyses leveren de onderstaande gegevens op:

$$F = \frac{(15288.109 - 14859.516) / (6-4)}{14859.516 / (77-6)} = 1.024,$$

$df_1 = 6-4=2$ en $df_2 = 77-6=71$.

De gevonden $F=1.024$ is niet significant, zodat er geen interactie tussen treatment en CBG is. Model II is derhalve een adequate descriptie van de data. De hypothese met betrekking tot de veronderstelde interactie moet verworpen worden.

Met behulp van variantie-analyses is ook nagegaan of het bovengeconstateerde beeld teruggevonden kon worden indien als afhankelijke variabele slechts die items van de natoets gebruikt zouden worden die betrekking hadden op doelstellingen van les 1, vervolgens op die van les 2, enz. In totaal derhalve vier variantie-analyses. Steeds bleek er sprake van een significant hoofdeffect CBG, terwijl het conditie-hoofdeffect en de interactie niet significant waren. Bij combinatie van de items voor de lessen 1 en 3 (de 'moeilijke' lessen) als afhankelijke variabele in een variantie-analyse kwam eveneens hetzelfde beeld naar voren. Slechts bij de combinatie van de items voor de lessen 2 en 4 (de 'gemakkelijke' lessen) als afhankelijke variabele in een variantie-analyse bleek de interactieterm significant ($p < .10$) te zijn. Gemiddelden en standaardafwijkingen voor deze twee combinaties zijn in tabel 8.8. vermeld.

Tabel 8.8. Gemiddelden en standaardafwijkingen op twee subtests van de Natoets

		'Moeilijke' leertaken			'Gemakkelijke' leertaken		
Conditie		I	II	III	IV	V	VI
Hoog-CBG	M	37.08	37.33	35.31	56.08	58.00	59.31
	SD	4.52	7.43	8.20	9.29	7.26	8.24
Laag-CBG	M	25.62	26.79	23.54	49.38	49.79	42.23
	SD	6.02	6.95	8.91	8.78	9.53	11.33

De gepresenteerde gegevens in tabel 8.8. maken twee dingen duidelijk. In de eerste plaats dat het onderscheid maken tussen totale en partiële herhaling als correctivum weinig zinvol is, noch voor leerlingen met een hoog, noch voor leerlingen met een laag cognitief begingedrag. In de tweede plaats kan geconstateerd worden dat het, afhankelijk van de moeilijkheidsgraad van de leertaak, verstrekken van nu eens totale dan weer partiële herhaling tot lagere resultaten leidt dan een keuze voor een van beide correctiva. Vooral leerlingen

met een (relatief) laag-CBG presteren in zo'n situatie minder.

Een variantie-analyse waarbij totale herhaling (condities I en IV) vergeleken werd met partiële herhaling (condities II en V) leverde geen significant verschil op.

De derde hypothese dient dus eveneens verworpen te worden. Bij de interpretatie van bovenstaande conclusies en data moet bedacht worden dat nog geen rekening is gehouden met eventueel optredende verschillen in benodigde leertijd. In de volgende paragraaf zal aan dit aspect aandacht besteed worden.

8.10.3. Toetsing der hypothesen na correctie voor leertijd

Tot nu toe is bij de toetsing van de diverse hypothesen uitsluitend uitgegaan van de absolute scores op na- en retentietoets zoals behaald door de proefpersonen in de diverse condities. Op analoge wijze als in het eerste experiment zullen wij ook nu deze scores corrigeren voor mogelijk optredende verschillen in leertijd. Zowel op de absolute scores van de na-, als van die op de retentietoets zullen efficiëntiescores (zie par. 7.5.2.1.) worden berekend. In tabel 8.9. worden deze descriptieve gegevens samengevat.

Tabel 8.9. Efficiëntiescores voor na- en retentietoets

Conditie		I	II	III	IV	V	VI
Eff.score (NT)	M	.715	.829	.700	.471	.488	.353
	SD	.141	.321	.254	.169	.157	.136
		(N=11)	(N=10)	(N=12)	(N=11)	(N=14)	(N=12)
Eff.score (RT)	M	.696	.822	.685	.469	.486	.358
	SD	.151	.307	.261	.152	.167	.152
		(N=11)	(N=10)	(N=12)	(N=11)	(N=15)	(N=12)

Uit de gegevens in tabel 8.9. zien we, in overeenstemming met het in het eerste experiment geconstateerde, dat de condities waarin partiële herhaling wordt toegepast (conditie II & V) efficiënter presteren dan condities waarin totale herhaling plaatsvindt (condities I en IV). Voor leerlingen met een hoog-CBG is deze trend duidelijker dan voor leerlingen met een laag-CBG. De afwisseling van totale en partiële herhaling is voor de hoog-CBG proefpersonen niet efficiënter in vergelijking met proefpersonen die uitsluitend totale

herhaling krijgen. Voor leerlingen met een laag-CBG is de afwisseling beduidend minder efficiënt.

Met behulp van variantie-analyses is wederom nagegaan of de condities significant van elkaar verschillen. De resultaten staan vermeld in tabel 8.10.

Tabel 8.10. Resultaten variantie-analyses op de efficiëntie-scores gebaseerd op na- en retentietoets.

Afh. var.	Bron	SS	df	MS	F	p
Eff.scores (NT) (N=70)	CBG	1.668	1	1.668	40.25	< .01
	Conditie	.208	2	.104	2.51	< .10
	Interacties	.038	2	.019	.45	ns
	Residu	2.653	64	.041		
Eff.scores (RT) (N=71)	CBG	1.537	1	1.538	37.04	< .01
	Conditie	.212	2	.106	2.55	< .10
	Interacties	.042	2	.021	.50	ns
	Residu	2.699	65	.042		

De resultaten van de variantie-analyses laten zien dat het hoofdeffect CBG significant is, zoals we ook reeds in de voorgaande paragraaf met betrekking tot de absolute scores geconstateerd hebben. Het conditie-hoofdeffect komt nu wat duidelijker naar voren dan bij de analyse van de absolute scores. Een en ander wordt veroorzaakt door de condities II en IV, zoals uit de gegevens in tabel 8.9. geconstateerd kan worden. De interactie tussen condities en CBG blijkt ook nu weer niet significant.

8.11. Discussie

Uit de resultaten van het in dit hoofdstuk beschreven experiment kunnen diverse conclusies getrokken worden. Leerlingen met een hoog cognitief begingedrag blijken steeds significant hogere leerprestaties op een summatieve toets te leveren dan leerlingen met een laag cognitief begingedrag. De beheersingslerenstrategie slaagt er dus niet in de invloed van deze initiële verschillen geheel uit te schakelen. Dit betekent dat het aanbeveling kan verdienen om lacunes in het kennisbestand van leerlingen met betrekking tot een bepaald leerstofdomein aan te vullen vóórdat deze leerlingen aan

het onderwijsleerproces deelnemen. Immers we hebben geconstateerd dat aanlegverschillen (gedeeltelijk) ondervangen kunnen worden door een goede instructieprocedure. Tekorten in voorkennis e.d. blijven echter doorwerken op latere leerprestaties.

Uit het onderzoek kwam ook naar voren dat totale herhaling (het Keller-correctivum) niet tot hogere leerprestaties leidt dan partiële herhaling (het Bloom-correctivum), noch voor proefpersonen met een hoog CBG, noch voor proefpersonen met een laag CBG. Deze bevinding is in overeenstemming met de in het eerste experiment gevonden resultaten en met de resultaten uit het onderzoek van Swanson (1977). De bevindingen weerspreken de resultaten van Tierney (zie: Block, 1974). Ook het maken van onderscheid tussen moeilijke en makkelijke leertaken in relatie tot de aard van de correctieve procedure levert geen significant verschil tussen totale en partiële herhaling op. De veronderstellingen met betrekking tot de aard van de door de hoog- en laag-CBG-proefpersonen gemaakte fouten worden derhalve niet door empirische gegevens ondersteund.

Het afhankelijk van de moeilijkheidsgraad van de leertaak nu eens toepassen van totale, dan weer van partiële herhaling leidt tot lagere leerprestaties dan het consequent toepassen van een der correctieve procedures. Wellicht worden de proefpersonen enigszins in verwarring gebracht door de verschillende procedures, waardoor zowel de leerprestaties als de benodigde leertijd in ongunstige zin worden beïnvloed.

Indien we rekening houden met de leertijd, dan blijkt dat partiële herhaling beter is dan totale herhaling, ongeacht het cognitief begingedrag. Dit effect kon zowel op de efficiëntiescores voor de na- als voor de retentietoets geconstateerd worden. Kennelijk vergt het bestuderen van een onderdeel van de leertaak in schemavorm minder tijd dan het overdoen van die gehele leertaak. De leerprestaties zijn overigens niet significant beter.

De conclusie kan derhalve luiden dat de aard van de correctieve procedure met betrekking tot de leerprestaties niet zo belangrijk is, alhoewel procedures die minder leertijd vergen wel bijdragen aan verhoging van de leerefficiëntie. Het afwisselen van correctieve procedures van verschillende aard leidt zowel tot lagere leerprestaties als tot vermindering van de leerefficiëntie.

Tenslotte kan nog opgemerkt worden dat voor alle proefpersonen de scores op de retentietoets nauwelijks afwijken van die op de natoets. Ook in het eerste experiment konden wij dit constateren, maar daar zou wellicht de oorzaak gelegen kunnen zijn in de bekendheid van de proefpersonen met de items daar deze in de voor- na- en retentietoets identiek waren. In dit

experiment is daarentegen gewerkt enerzijds met gedeeltelijk nieuwe items en anderzijds met gewijzigde item-volgordes. Ook nu kwam naar voren dat de retentie-toetsscores nauwelijks lager waren dan de natoetsscores.

ALGEMENE DISCUSSIE

9.1. De belangrijkste onderzoeksresultaten

In hoofdstuk 7 van deze studie is de vraag of de strategie van beheersingsleren, zoals door ons in de cursus geoperationaliseerd, zich in de praktijk wel zo manifesteerde als in theorie verondersteld werd, bevestigend beantwoord. Uit de resultaten van het eerste experiment bleek onder meer dat de leerresultaten na beëindiging van een leertaak door toepassing van feedback en correctie veranderbaar waren. Bovendien constateerden we dat de variantie in leerprestaties in de groep leerlingen die volgens de methode van het beheersingsleren was onderwezen geringer was, dan in de groep leerlingen waarbij dit niet het geval was.

Uit de experimentele resultaten konden we afleiden dat de correlatie tussen aanleg ('aptitude') en leerprestaties voor de leerlingen uit de beheersingslerenconditie lager was dan die voor de andere leerlingen. Weliswaar was de grootte van de gevonden correlatie niet geheel in overeenstemming met hetgeen door het model voorspeld werd, de trend die geconstateerd kon worden, was daarentegen overtuigend genoeg.

Vervolgens werd nagegaan of de diverse componenten uit het model van Bloom dat percentage variantie in leerlingprestaties konden verklaren als door het model werd gepostuleerd. Met betrekking tot die componenten werd het percentage verklaarde variantie berekend. Eenmaal zonder en eenmaal met inachtneming van de verschillen in leertijd tussen de condities als gevolg van de toepassing van feedback en correctieve procedures.

De variabele cognitief begingedrag (CBG), bestaande uit een toets voor de meting van de voorkennis en een aptitudetest, verklaarde in de beheersingslerenconditie 46.6% en in de non-mastery-conditie 62.2% van de variantie op een leerdoelgebonden natoets. Na correctie voor de verschillen in leertijd bedroegen deze percentages resp. 39.3% en 67.8%. De modelvoorspelling ongeacht de condities was 50%. Zowel binnen de mastery- als de non-mastery-conditie bleek voorkennis een grotere variantieverklarende waarde te hebben dan aptitude. Na correctie voor tijd was de variantieverklarende waarde van de aptitude-variabele in de mastery-conditie zeer gering en nam het belang van de voorkennis toe. Dit door ons gevonden belang van de voorkennis als variantieverklarende variabele is in overeenstemming met diverse andere theoretische veronderstellingen en empirische bevindingen. Tobias (1976) wijst op de variabele 'prior achievement' als zijnde een uiterst

essentiële variabele in het ATI-onderzoek. Het begrip 'aptitude' in dit verband wordt door hem zeer ruim opgevat, aangezien de term niet bij uitsluiting gereserveerd wordt voor de meer persoonsgebonden eigenschappen. Ook Cronbach & Snow (1977) gaan uit van een dergelijk algemeen aptitude-begrip. Lodewijks (1981) wijst op het belang van de voorkennis als variantieverklarende variabele voor leerprestaties gemeten met een summatieve toets. Hij rapporteert dat voorkennis plm. 36% van de variantie in leerprestaties op een natoets kan verklaren (o.c., p. 273). Hij heeft bij de vaststelling van het niveau van voorkennis overigens geen gebruik gemaakt van een voortoets, maar van het subjectieve oordeel van de leerling over de mate waarin de leerling vertrouwd is met de aan te bieden leerstof. Deze aanpak sluit aan bij Tobias' (1976) begrip 'familiarity'. Uit het door Lodewijks (1981, p.233-278) uitgevoerde experiment blijkt dat hoe meer een leerling van mening is dat de leerstof hem bekend voorkomt, des te hoger zijn leerprestaties zijn.

Ook Schmid-Schönbein (1980) wijst op het belang van de voorkennis bij de remediële instructie met betrekking tot het maken van fouten bij het leren van Engels als een tweede taal. Het belang van de voorkennis als variantieverklarende variabele voor latere leerprestaties heeft mede aanleiding gegeven tot het tweede door ons uitgevoerde experiment. De veronderstelling daarbij was dat leerlingen met een hoog CBG een minder omvattende vorm van feedback en correctie nodig zouden hebben dan leerlingen met een relatief laag CBG. Dit uitgangspunt sluit aan bij de opvatting van Tobias (1976, p.72) dat leerlingen met een hoog cognitief begingedrag ('prior achievement') aan een instructieprocedure toegewezen kunnen worden waarin hen minimale ondersteuning wordt verschaft, terwijl leerlingen met een laag cognitief begingedrag juist maximale ondersteuning in het onderwijsleerproces behoeven. Op basis van het door ons uitgevoerde experiment kon niet geconcludeerd worden tot een interactie tussen het cognitief begingedrag en de aard van de toegepaste correctieve procedure; noch met, noch zonder correctie voor mogelijke verschillen in leertijd. Deze bevinding is in overeenstemming met hetgeen Lodewijks (1981, p.258-260) in dit verband heeft gevonden. Ook hij kon geen interactie tussen voorkennis (gemeten door middel van het subjectieve oordeel van de leerling) en condities vaststellen, wel een duidelijk hoofdeffect.

De volgende variabele in het model van Bloom die wij op zijn variantieverklarende waarde onderzocht hebben, heeft betrekking op de affectieve leerlingkenmerken.

Het percentage variantie in leerlingprestaties op een

summatieve natoets dat door deze variabele kon worden verklaard, was zowel met als zonder correctie voor leertijdverschillen, nihil. Het model voorspelde daarentegen plm. 25%. Deze discrepantie is ons inziens zodanig dat men zich kan afvragen of het model op dit aspect geen bijstelling behoeft. Block & Burns (1977) stellen naar aanleiding van hun overzicht van empirische bevindingen dat de empirische data weliswaar in de voorspelde richting wijzen, doch dat een mogelijk Hawthorne-effect niet is uitgesloten (zie ook: Hoofdstuk 2). Ook zal, duidelijker dan tot nu toe is gebeurd en ook duidelijker dan in deze studie is geschied, aangegeven moeten worden welke de relatie is tussen de affectieve aanvangskenmerken van de leerlingen en de latere cognitieve leerprestaties. Met name ten aanzien van dit aspect is de ontwikkeling van een miniatuurtheorie geïndiceerd.

De derde variabele in het model tenslotte is de 'kwaliteit van de instructie'. De in ons onderzoek aangebrachte variaties met betrekking tot deze variabele hebben alle te maken met de component feedback en correctie. Het percentage variantie dat door de variabele kwaliteit van de instructie kon worden verklaard, bedroeg op de summatieve natoets 8.2%. Indien we rekening hielden met de variantie in feedbackprocedures dan steeg dit percentage tot 9.2%. Echter, bij in aanmerking nemen van de leertijd daalde dit percentage tot slechts 1%, terwijl de bijbehorende F-ratio bovendien niet significant was. Het model voorspelt een percentage verklaarde variantie van 25; onze bevindingen wijken daar nogal vanaf. Ook Nuy (1981, p.215 e.v.) vond een met onze bevindingen vergelijkbaar percentage. De instructievariabele in de vergelijking van een mastery-learning conditie en een traditionele onderwijsconditie bleek in staat 7.96% van de variantie in leerlingprestaties op een natoets te verklaren, waarbij wel reeds de algemene voorkennis was uitgepartialiseerd, maar geen rekening was gehouden met verschillen in leertijd.

Als we vervolgens ook de resultaten van het onderzoek naar de verschillende feedbackprocedures in ogenschouw nemen, dan blijkt dat feedback en correctie belangrijke aspecten zijn in het onderwijsleerproces. Er kan niet gesteld worden dat de Keller-correctieve procedure beter of slechter is dan de Bloom-correctieve procedure. Deze bevinding is in overeenstemming met de resultaten uit het onderzoek van Tierney (1973). Indien we rekening houden met de leertijd dan is er enige indicatie dat de Bloom-procedure efficiënter is dan de Keller-procedure. Van enige interactie met de moeilijkheidsgraad van de leerstof, of met het cognitief begingedrag van de leerlingen is in ons onderzoek niets gebleken.

Resumerend kan gesteld worden dat het model van Bloom wellicht bijstelling behoeft aangaande de rol van de affectieve leerlingkenmerken, terwijl voorts ook de percentages verklaarde variantie bijgesteld moeten worden. Bovendien verdient het aanbeveling consequent de invloed van de leertijd in het model te verdisconteren. Verschillende percentages kunnen daardoor een ander gezicht en gewicht krijgen.

Tot nu toe is nauwelijks gesproken over de aard van onze steekproef. De homogeniteit van de proefgroep kan mogelijk-
wijs ook van invloed zijn op de percentages verklaarde variantie. Alhoewel het model algemene geldigheid pretendeert, is het toch zinvol om rekening te houden met de verschillende soorten populaties waarnaar men bepaalde empirische bevindingen wenst te generaliseren. De homogeniteit of heterogeniteit van de populatie behoort in het model een plaats te krijgen.

9.2. Theoretische implicaties

9.2.1. Inleiding

In het navolgende wordt ingegaan op de relatie tussen de beheersingsleren-strategie en de ATI-benadering. Wij sluiten daarbij aan bij een discussie die einde 1979 in het Tijdschrift voor Onderwijsresearch is gevoerd. Tevens wordt gepoogd de opvattingen betreffende het beheersingsleren in een ruimer theoretisch kader te plaatsen, en wel door aansluiting te zoeken bij de theorievorming betreffende 'direct instruction', of 'active teaching' (zie ook hoofdstuk 3, par. 3.4.2. van deze dissertatie). Met betrekking tot dit laatste aspect is wellicht enige toelichting gewenst. Immers in ons onderzoek is gebruik gemaakt van een 'leraarloze' cursus, terwijl de theorievorming betreffende het 'active teaching model' wel degelijk te maken heeft met leerkrachtgedrag.

Wij menen dat in de geïndividualiseerde cursus zoals in dit onderzoek gebruikt, diverse taken die normaliter door de leerkracht uitgevoerd worden nu *in* het cursus-materiaal opgenomen zijn, terwijl de beslissingen aangaande voortgang en correctieve procedures nu tot de taken van de proefleider behoren. Het onderzoek met betrekking tot het leerkrachtgedrag kan, althans tot op zekere hoogte, wel degelijk van belang zijn voor de door ons gekozen onderwijs-vorm binnen de strategie van beheersingsleren, terwijl deze voor de strategie *sec* zonder meer van belang kan zijn.

9.2.2. *Beheersingsleren en ATI-onderzoek*

In ons tweede experiment hebben wij binnen de strategie van beheersingsleren een 'aptitude-treatment-interactie (ATI)-onderzoek' uitgevoerd. Als aptitude-variabele fungeerde het cognitief begingedrag en als treatment-variabele werden twee correctieve-procedures toegepast. Er kon geen interactie aangetoond worden. In het licht van het tot nu toe verrichte ATI-onderzoek is dit zeker geen onverwacht resultaat. Sedert het verschijnen van het inmiddels klassieke artikel van Cronbach (1957) betreffende het ATI-paradigma als integratie van tot dan twee afzonderlijke disciplines, de experimentele en de correlatieve psychologie, heeft er heel wat ATI-onderzoek plaatsgevonden. Het resultaat is teleurstellend: vrijwel geen gerepliceerde interacties die aangeven dat de ene onderwijsmethode ideaal is voor leerlingen met kenmerk x, terwijl de andere onderwijsmethode optimaal is voor leerlingen met kenmerk y. De vraag is nu waarom zo'n intuïtief alleszins plausibel paradigma zo weinig concrete en bruikbare resultaten oplevert. Voor de beantwoording van deze vraag moeten we eerst de twee uitgangspunten van het ATI-paradigma nader beschouwen. Het eerste uitgangspunt is dat leerlingen met betrekking tot aptitudes verschillen, waarbij deze verschillen voor de resultaten van onderwijsleerprocessen van belang zijn. Het tweede uitgangspunt is, dat onderwijs-arrangementen verschillende leerresultaten tot gevolg kunnen hebben.

De individuele verschillen waarover het gaat betreffen verschillen in aptitudes. Een 'aptitude' is in dit verband een complex van persoonlijke kenmerken, welke verklaringswaarde hebben voor het eindgedrag dat een leerling vertoont na een bepaalde instructiewijze of 'treatment' (Weeda, 1973). Snow (1977b) definieert 'aptitude' als een individuele verschilvariabele die leerresultaten in een bepaalde situatie voorspelt. Crombag (1979) merkt op, dat deze volledig operationele definitie slechts aangeeft hoe achteraf (post hoc) kan worden bepaald wat een aptitude is, namelijk alles wat leerresultaten blijkt te kunnen voorspellen. Alhoewel deze kritiek iets te ver gaat, wordt daardoor toch wel een kernprobleem van het ATI-paradigma aangeduid, namelijk dat het niet tevoren op grond van een theorie evident is welke variabelen verklaringswaarde hebben met betrekking tot leerprestaties. De keuze van de aptitude-variabele blijft daardoor enigszins willekeurig.

Voor Crombag (1979, p. 180) is het bovendien nog een vraag of de verschillen in leerresultaten werkelijke verschillen in vaardigheden reflecteren of slechts door de constructiewijze van het meetinstrument gecreëerd worden. Hij ziet in een en

ander aanleiding om de theorie van het beheersingsleren, waarin wordt aangegeven dat het onderwijsarrangement de individuele verschillen in leerresultaten juist moet minimaliseren, boven het ATI-paradigma te prefereren. Deze preferentie is bekritiseerd door De Klerk (1979) en Verhoeven (1979).

De Klerk (1979, p. 193), en ook Verhoeven (1979, p. 195), formuleren als bezwaar tegen de opvatting van Crombag dat het ATI-paradigma een onderzoeksmethode is, terwijl de mastery learning-strategie een onderwijsmethode is. Onvergelijkbare grootheden derhalve. Wij menen dat deze kritiek in dit geval op een misverstand berust (waar de door Crombag gebruikte formulering overigens aanleiding toe geeft). Crombag verwijst bij zijn keuze voor het beheersingslerenmodel vooral naar de publikatie van Bloom betreffende het model voor het leren op school (Bloom, 1976). (De publikatie van Block (1971) waarin de *onderwijs*methode beheersingsleren wordt beschreven, doet Crombag als naïef af.) Dit model (zie Hoofdstuk 3 van dit proefschrift) geeft aan welke variabelen een rol spelen bij de totstandkoming van leerresultaten. Hierdoor kan het model hypothesen-genererend werken en als onderzoeksparadigma fungeren. Als zodanig kan het, naar de mening van Crombag, als alternatief voor het ATI-paradigma gelden.

Het tweede uitgangspunt van het ATI-paradigma is, dat verschillen in onderwijsmethoden of -arrangementen verschillen in leerresultaten van leerlingen tot gevolg hebben. Crombag (1979, p. 181) merkt op, dat de empirische evidentie die dit uitgangspunt moet ondersteunen erg gering is. Als verklaring oppert hij (o.c., p. 182) dat de vergeleken onderwijsmethoden te weinig van elkaar verschillen. Crombag beveelt de Keller-plan-strategie aan als onderwijsmethode die voor alle leerlingen, ongeacht hun aptitudes, de beste onderwijsmethode is. Deze vorm van beheersingsleren sluit aan bij de voorkennis van de leerling en niet bij zijn aptitude-profiel.

Crombag meent, ons inziens terecht, dat de Keller-plan-strategie past binnen het theoretisch model van het beheersingsleren (Bloom, 1976), zodat ook vanuit theoretisch gezichtspunt geen behoefte is aan gedifferentieerde ATI-modellen. De Klerk (1979, p. 192-194) en Verhoeven (1979, p. 197) bestrijden de opvatting van Crombag aangaande de geringe theoretische achtergrond van het ATI-paradigma. De Klerk (o.c., p. 194) stelt dat dit bezwaar a fortiori voor het beheersingsleren geldt. Het bezwaar van Crombag wordt daardoor echter niet minder waar! Overigens wordt door Bloom (1976) ons inziens in voldoende mate gepoogd zijn model theoretisch en empirisch te onderbouwen. Van andere orde is De Klerk's aanbeveling om met betrekking tot de theorievorming aan te

sluiten bij de benaderingswijze van Hunt & Sullivan (1974). Zij stellen voor leerlingkenmerken te definiëren in termen van 'accessibility characteristics': leerlingkenmerken die rechtstreeks verwijzen naar specifieke onderwijsmaatregelen, of instructiemethoden. Een soortgelijke aanbeveling wordt gedaan door Verhoeven (1979, p. 197) die voorstelt om bij de keuze van aptitudevariabelen uit te gaan van de treatments. Zowel De Klerk als Verhoeven zien vele voordelen in het afstemmen van onderwijsmethoden op aptitudes ('matching'). Wij vragen ons af of de onderwijspraktijk ook gediend is met de (eventuele) resultaten van dergelijk onderzoek. Immers, voor het ene vak is een bepaalde aptitude van belang, voor het andere vak weer een andere. Voor het ene leerjaar zijn het deze aptitudes, voor een volgend leerjaar zijn het weer andere die de onderwijsmethode bepalen. Bovendien kan dit binnen een vak en binnen een leerjaar ook nog variëren, terwijl het zeker over leerlingen zal variëren. Hoe dit alles in de dagelijkse onderwijspraktijk geïmplementeerd moet (of: kan?) worden blijft onduidelijk.

In dit verband geven wij de voorkeur aan de strategie voor beheersingsleren, alsmede aan de theoretische fundering daarvan door Bloom (1976). Wel menen wij mogelijkheden te zien resultaten van ATI-onderzoek, zo deze replicerbaar blijken, toe te passen binnen de mastery learning-benadering. Intussen moeten ook wij echter concluderen dat het door ons uitgevoerde ATI-onderzoek geen zinvolle interacties heeft opgeleverd, en dat de vergelijking van correctieve procedures niet geleid heeft tot een duidelijke keuze voor of de Bloom- of de Kellerstrategie. Een verklaring hiervoor, alsmede voor het tot nu toe teleurstellende ATI-onderzoek, kan wellicht gevonden worden in het construct *macro-processen*.

In een recent artikel van Tobias (1982) wordt de hypothese geformuleerd dat onderzoek gericht op vergelijking van de externe kenmerken van instructiemethoden voorbij gaat aan de belangrijkste verklarende variabele van leerresultaten, namelijk *macro-processen*: de frequentie en intensiteit waarmee leerlingen cognitief de instructie-input verwerken. Als verschillende instructiemethoden vergelijkbare resultaten tot gevolg hebben, dan vinden bij de leerlingen waarschijnlijk dezelfde macro-processen plaats.

Als door een bepaalde instructieprocedure de leerprestaties van leerlingen worden verhoogd, dan kan dit succes waarschijnlijk toegeschreven worden aan het feit dat het de leerlingen heeft gestimuleerd actiever met de leerstof om te gaan teneinde deze te begrijpen. Volgens de macro-proces-hypothese zal elk onderwijs-arrangement dat leidt tot een

actievere en intensievere verwerking van de instructie-input door de leerling, leiden tot verbetering van de leerprestaties. Als we deze opvatting betreffende macro-processen relateren aan het ATI-paradigma, dan betekent dit dat we niet simpelweg instructiemethoden moeten variëren en vervolgens kijken of dit invloed heeft op de leerresultaten, maar dat we die instructiemethoden moeten selecteren die verschillende macro-processen induceren. Koppelen we deze opvatting terug naar ons tweede experiment waarin we twee correctieve procedures vergeleken, dan kan het ontbreken van een significant verschil in leerprestaties wellicht toegeschreven worden aan het feit dat beide correctiva waarschijnlijk dezelfde macro-processen geïnduceerd hebben.

9.2.3. Beheersingsleren en het 'direct teaching model'

In de voorgaande paragraaf is verondersteld dat het onderwijsarrangement, dat is opgezet conform de principes van de beheersingslerenstrategie, een macro-proces bij de leerlingen induceert, waardoor zij zeer actief 'omgaan' met de leerstof. Dit proces wordt geïnduceerd door de strikte opbouw van de leertaak, de frequente toetsingsmomenten en de zeer gerichte correctieve procedures. Hierdoor ontstaat een efficiënt leerproces, waarbij een groot gedeelte van de door de leerling gebruikte leertijd ('time spent') ook werkelijk taakgerichte leertijd is ('time on task'), dat wil zeggen gebruikt wordt om de leerdoelen van deze specifieke leertaak te bereiken. (Terzijde zij opgemerkt dat niet alleen de hoeveelheid beschikbare leertijd - 'time allowed' - van belang is, maar vooral wat er in die leertijd door de leerling wordt gedaan). In hoofdstuk 3 (par. 3.4.2.) hebben wij gewezen op een vijftal benaderingen (waar onder het beheersingsleren) in het onderzoek betreffende de leertijd. Steeds kwam daarbij naar voren dat effectieve vormgeving van het onderwijsleerproces de taakgerichte leertijd doet toenemen. Deze opvatting, waarvoor in dit proefschrift enige empirische evidentie is aangedragen, sluit aan bij theorieën en modellen aangaande het onderwijsleerproces die de laatste jaren aangeduid zijn met de term 'direct instruction'. Hieronder worden onderwijsactiviteiten verstaan, waarbij leerlingen de leerdoelen helder voor ogen hebben, waar voldoende tijd beschikbaar wordt gesteld en waar veel leerinhouden aan bod komen. Op de verrichtingen van de leerlingen wordt nauwlettend toegezien. De gegeven terugkoppeling is onmiddellijk en inhoudelijk gericht op het gegeven antwoord. Bij directe instructie bepaalt de leraar (c.q. de leergangconstructeur) de leerdoelen, hij kiest de

voor de leerling geschikte leermaterialen (cf. Veenman, 1980, p. 210-211).

In bovenstaande omschrijving, maar ook in de definities van anderen (o.a. Bennett, 1976, p. 17-18; Good, 1979, p. 55; Rosenshine, 1979, p. 38) betreffende 'direct instruction', of 'active teaching' wordt in sterke mate de nadruk gelegd op de rol van de leerkracht als 'manager van het onderwijsleerproces' (vergelijk de opvatting van Keller (1968) aangaande de rol van de leerkracht; zie ook par. 2.4. van dit proefschrift). Deze zogenaamde 'monitoring'-functie van de leerkracht vinden we ook terug in de beheersingslerenstrategie. Het is ook deze functie die verantwoordelijk geacht kan worden voor de inductie van het in de voorgaande paragraaf bedoelde macro-proces.

Het onderzoek betreffende de directe instructie is, evenals de mastery-learning benadering, gebaseerd op de ideeën van Carroll (1963) en bovendien op het Wiley-Harnischfeger model (1974). De variabelen die onderscheiden worden (gedefinieerd in termen van tijd) zijn tot op zekere hoogte controleerbaar. De leerkracht kan leerprestaties van leerlingen beïnvloeden door bepaalde onderwijsproblemen te formuleren in termen van variabelen waarop hij enige greep heeft. Block (1980) noemt dit de 'we can do'-benadering, welke hij plaatst tegenover de 'we can't do'-benadering. In de laatste benadering wordt de oplossing van problemen waarvoor het onderwijs staat, steeds geformuleerd in termen van exogene factoren waar de mensen in het onderwijs geen greep op hebben (Veenman, 1980, p. 281).

De theorievorming betreffende de beheersingslerenstrategie kan aansluiten bij die betreffende de directe instructie, daar beide vergelijkbare macro-processen oproepen.

9.3. Praktische implicaties

Het primaire doel van het in deze studie beschreven onderzoek is het leveren van een bijdrage aan de wetenschappelijke fundering en uitbouw van een onderwijskundig verklaringsmodel dat ten grondslag ligt aan de strategie van beheersingsleren. Dit neemt niet weg dat er ook enige conclusies uit het onderzoek te trekken zijn die meer praktische betekenis kunnen hebben. Gebleken is in ieder geval dat het toepassen van de principes van het beheersingsleren tot hogere leerprestaties leidt dan wanneer dergelijke principes niet toegepast worden. Bovendien is gebleken dat de retentie der leerresultaten alleszins redelijk is. Ten aanzien van de aard der correctieve procedure kan gesteld worden dat er enige evidentie bestaat dat het toepassen van de een of andere vorm van correctie

beter is dan het nalaten van correctie, zowel voor goede als voor minder goede leerlingen.

Er is enige reden te veronderstellen dat partiële remediële instructie (het Bloom-type correctivum) efficiënter werkt dan totale herhaling (het Keller-type correctivum). Geen rekening is gehouden met het feit dat het *construeren* van partiële remediële instructie-procedures ook tijd kost. Indien men die tijd niet heeft, of niet daarvoor wenst te besteden, dan is er weinig bezwaar om te volstaan met eenvoudige herhaling van de leertaak. Aangenomen dat deze taak niet te omvangrijk is.

Bij het bovenstaande moet men overigens wel in aanmerking nemen dat de empirische evidenties die aan een en ander ten grondslag liggen gebaseerd zijn op een tamelijk homogene populatie studenten hoger onderwijs, op een cursus die voornamelijk individueel doorgewerkt kan worden en op leerstof waarmee de leerlingen al reeds enige mate vertrouwd zijn en waarvan ze ook al reeds enige voorkennis bezitten.

9.4. Suggesties voor verder onderzoek

Reeds diverse malen is het grote belang van het cognitief begingedrag aangegeven als verklarende variabele voor de variantie in leerprestaties op een summatieve toets. Dit CBG zoals wij dit geoperationaliseerd hebben, bestond uit twee variabelen: de voorkennis en de aanleg ('aptitude').

Met betrekking tot de voorkennis bestaan er nog diverse onduidelijkheden aangaande de wijze waarop deze variabele zijn invloed in het onderwijsleerproces doet gelden. Leertaak-analyses en metingen van de voorkennis kunnen wellicht hier wat meer helderheid in brengen. Voor wat betreft de meting van de voorkennis is het goed een onderscheid te maken tussen leerstof waarvan de leerlingen al reeds een zekere mate van voorkennis hebben en leerstof waarbij dat slechts in zeer geringe mate het geval is. Voor de eerste categorie leerstof is het gebruik van een voortoets het meest geïndiceerd. Voor de tweede categorie kan wellicht gebruik gemaakt worden van de meting der subjectieve competenties van leerlingen (Lodewijks, 1981). Hierbij wordt enerzijds nagegaan de mate waarin de leerling van mening is dat de te onderwijzen leerstof hem bekend voorkomt en anderzijds wordt nagegaan de mate waarin hij die stof als moeilijk of gemakkelijk beoordeelt. Ter aanvulling van de gebruikelijke voortoetsscores kunnen deze gegevens wellicht een nuttige functie vervullen. Wij menen niet dat deze de voortoets volledig kunnen vervangen, terwijl voorts de bruikbaarheid beperkt blijft tot het onderzoek.

Bovendien biedt de meting ook weinig specifieke informatie aangaande lacunes in de kennis. Als aanvulling op het onderzoeksinstrumentarium kan het echter wel van nut zijn.

Wij menen dat het onderzoek met betrekking tot interne differentiatie in het onderwijs zich ook bij voorkeur dient te richten op de voorkennis als ingang van differentiatie. Wij verwachten van dergelijke dicht bij het onderwijsleerproces en de leerstof staande variabelen meer rendement dan van allerlei psychologische variabelen die te weinig rechtstreeks gekoppeld zijn aan het concrete onderwijsgebeuren. Doordat deze variabelen generaliseerbaar moeten zijn over verschillende vakken, schooltypen, leeftijden en dergelijke worden ze noodgedwongen veelal te algemeen en daardoor onvoldoende geschikt voor differentiatie. In dit verband kan een parallel getrokken worden met de ten aanzien van de leerprestaties afnemende predictieve waarde van het IQ ten opzichte van meer leerstofgebonden toetsen (cf. Bloom, 1976).

De tweede variabele die van belang is in het cognitieve begin-gedrag is de 'aptitude'. De aptitude-meting zoals wij die in deze studie toegepast hebben, is in feite nog klassiek. We hebben middels de daarvoor bestaande instrumenten in feite de potentie tot bepaalde leerprestaties door de leerlingen gemeten. Bloom daarentegen is, in navolging van Carroll, er een voorstander van aptitude te definiëren in termen van tijd nodig om een bepaald criterium te behalen. Dit tijdsaspect van de aptitude-variabele is nog tamelijk ondoorzichtig. In het tweede hoofdstuk van dit proefschrift hebben wij er reeds op gewezen, dat Cronbach & Snow (1977) de mogelijkheid opperden dat de leertijden van leerlingen een multivariaat karakter zouden kunnen hebben. In een onderzoek van Glasnapp, Deaton & Soholove (1978) wordt enige empirische evidentie aangedragen voor de notie dat aptitude niet zonder meer gedefinieerd kan worden als de leertijd benodigd om een bepaalde criteriumscore te bereiken. Nader onderzoek met betrekking tot dit aspect is gewenst.

SAMENVATTING

Beheersingsleren: het model getoetst in de tijd

Het doel van dit proefschrift is de empirische toetsing van een door Bloom (1976) voorgesteld model betreffende het leren op school. Met name wordt aandacht geschonken aan de variantie in leerlingprestaties op een summatieve toets die door de diverse modelvariabelen wordt verklaard. Het onderwijsleerproces is gestructureerd volgens de principes van de strategie voor beheersingsleren.

Het proefschrift is globaal in drie delen opgebouwd.

Het eerste deel, omvattend de hoofdstukken 1 t/m 4, is een theoretisch deel waarin het model van Bloom uiteengezet wordt en aandacht wordt besteed aan individuele verschillen tussen leerlingen. Tevens bevat dit gedeelte van het proefschrift de vraagstellingen van het onderzoek.

Het tweede deel, de hoofdstukken 5 en 6, geeft een beschrijving van de in ons onderzoek gebruikte cursus en de daarbij geconstrueerde formatieve toetsen.

Het derde deel tenslotte, omvattend de hoofdstukken 7, 8 en 9, is het onderzoeksgedeelte waarin een tweetal experimenten wordt beschreven.

Hoofdstuk 1 bevat een korte uiteenzetting betreffende de relatie tussen onderwijspsychologie en onderwijskunde. Voorts wordt ingegaan op de ontwikkeling van miniatuurtheorieën aangaande (deel)aspecten van onderwijsleersituaties.

Hoofdstuk 2 bevat uiteenzettingen over individuele verschillen tussen leerlingen en over onderwijskundige afstemmingsmodellen. Aangegeven wordt hoe het onderwijs door middel van variëren van leerdoelen, instructievormen of beide afgestemd kan worden op individuele verschillen tussen leerlingen.

In dit hoofdstuk wordt tevens aandacht besteed aan het onderwijsleermodel van Carroll (1963), waarin verschillen tussen leerlingen gedefinieerd worden in termen van tijd. De verschillen in leerprestaties tussen leerlingen zijn niet alleen afhankelijk van de karakteristieken van individuele leerlingen, maar ook van de instructiekenmerken, c.q. de kwaliteit van de instructie.

De door het Carroll-model geïnspireerde *strategie van beheersingsleren* (Bloom, 1968) weerspiegelt een opvatting aangaande de optimale vormgeving van het onderwijsleerproces. De strategie wordt gekenmerkt door minimaal vijf aspecten, t.w.:

1. de doelstellingen van een cursus liggen vast;
2. de leerstof is opgedeeld in leertaken; leertaak k moet

beheerst worden alvorens aan leertaak (k+1) wordt begonnen;

3. na elke leertaak wordt een toets afgenomen en feedback verschaft;

4. bij onvoldoende beheersing van de leertaak wordt een correctivum aangeboden;

5. er is tempodifferentiatie, maar geen niveaudifferentiatie.

De beheersingslerenstrategie houdt rekening met individuele verschillen door tempodifferentiatie toe te passen, terwijl voor leerlingen die de leerstof na afloop van het (uniforme) instructieproces nog niet beheersen eventueel verschillende correctieve maatregelen genomen kunnen worden.

Tenslotte wordt in dit hoofdstuk het Keller-plan besproken - een vorm van beheersingsleren - en wordt ingegaan op enig empirisch onderzoek betreffende de strategie van beheersingsleren.

Hoofdstuk 3 geeft een overzicht van het model van Bloom (1976). Het model is bedoeld als verklaringsmodel voor verschillen in leerresultaten bij leerlingen.

Aan de input-kant bevinden zich de leerlingkenmerken: het cognitief begingedrag en de affectieve leerlingkenmerken. Het cognitief begingedrag kan plm. 50% van de variantie in leerlingprestaties op de summatieve toets verklaren. De affectieve leerlingkenmerken hebben betrekking op de attitude en de interesse van de leerlingen om aan bepaalde leertaken te beginnen. Deze kenmerken kunnen plm. 25% van de variantie in leerlingprestaties op een summatieve toets verklaren. In het procesgedeelte van het model is als belangrijkste variabele de kwaliteit van de instructie opgenomen. Een kwalitatief goed instructieproces is dat, wat voldoet aan de principes van het beheersingsleren. Deze variabele kan plm. 20-25% van de variantie in leerlingprestaties op de summatieve toets verklaren. Het totale model verklaart plm. 90% van de variantie in summatieve toetsprestaties.

De output-kant van het model betreft de cognitieve leerprestaties, de affectieve leerresultaten en de leertijd.

In het hoofdstuk wordt voorts aandacht besteed aan diverse onderdelen van het model afzonderlijk aandacht besteed, onder meer aan feedback en correctieve procedures en aan de leertijd. Tevens worden kritische kanttekeningen geplaatst bij de door het model verklaarde percentages variantie.

Het eerste deel van dit proefschrift wordt afgesloten met *Hoofdstuk 4*, waarin de vraagstellingen meer in detail uitgewerkt zijn. Het uitgangspunt van de vraagstellingen wordt gevormd door de vraag in hoeverre de variabelen uit het model van Bloom in een concrete situatie in staat zijn variantie in

leerlingprestaties te verklaren. Tevens zijn vraagstellingen geformuleerd betreffende de veronderstelde effecten van de beheersingslerenstrategie, onder meer in hoeverre toepassing van de strategie ertoe leidt dat het verband tussen aanleg en leerprestaties afneemt. Bovendien wordt rekening gehouden met de leertijd die een alternatieve verklaring kan vormen voor mogelijke verschillen in leerprestaties.

Het tweede deel van het proefschrift beschrijft de in het onderzoek gebruikte cursus en de formatieve toetsen. In *Hoofdstuk 5* wordt een descriptie gegeven van de cursus Frans, bestemd voor studenten hoger onderwijs die reeds enige voorkennis bezitten. De cursus omvat een zevental lessen over grammaticale problemen. Het geheel kan individueel en in eigen tempo doorgewerkt worden. Elke les wordt afgesloten met een formatieve toets.

Hoofdstuk 6 bevat de psychometrische analyse van de formatieve toetsen. De toetsen kunnen beschouwd worden als criterium-georiënteerde (leerstofgebonden) toetsen.

Het hoofdstuk bestaat uit twee delen. In het eerste gedeelte wordt aandacht besteed aan de bepaling van aftestgrenzen, alsmede aan de consistentiebepaling van de leerstofgebonden toetsen. Het tweede gedeelte van het hoofdstuk is gewijd aan de concrete toepassing van de technieken op de door ons gebruikte toetsen. Voor de bepaling van de aftestgrenzen wordt een drietal methoden gehanteerd:

1. de intuïtieve 80% methode: de proefpersonen moeten minimaal 80% van de items goed scoren (d.i. de methode volgens welke bij de uitvoering van het experiment te werk gegaan is);
2. De Bayes-procedure: die aftestgrens wordt gekozen waardoor de gemiddelde utiliteit van alle te nemen beslissingen zo groot mogelijk is;
3. de maximin-procedure: die aftestgrens wordt gekozen waardoor de utiliteit van de beslissing met de kleinste utiliteit van allemaal, nog zo groot mogelijk is.

Voor de o.g.v. de eerste methode bepaalde aftestgrenzen is ter vergelijking ook de Bayes- en maximin-utiliteit berekend. In het algemeen bleken de utiliteiten voor de diverse aftestgrenzen op de formatieve toetsen uit ons onderzoek voldoende hoog te zijn.

Bij de consistentiebepaling is een tweetal methoden toegepast, t.w.:

1. de methode van Huynh (1976), waarbij coëfficiënt κ wordt berekend;
2. de methode van Subkoviak (1976), waarbij coëfficiënt p_0 wordt berekend.

De consistentiebepaling geschiedt voor alle drie der bepaalde

aftestgrenzen. Over het algemeen bleken de consistentie-coëfficiënten voldoende groot te zijn.

Het derde deel van het proefschrift tenslotte beschrijft het uitgevoerde empirisch onderzoek. Het in *Hoofdstuk 7* beschreven experiment is gericht op de systematische toetsing van het model van Bloom. Tevens is aandacht besteed aan enkele verschillende vormen van feedback en correctieve procedures.

Twee condities zijn onderscheiden: een conditie die onderwijs krijgt gestructureerd volgens de principes van het beheersingsleren en een conditie waarbij dit niet het geval is. Het onderscheidend element tussen beide condities wordt gevormd door de aan- of afwezigheid van feedback en correctieve maatregelen. Binnen de beheersingsleren-conditie worden twee typen correctiva onderscheiden: een Bloom-type (partiële herhaling met behulp van een schema) en een Keller-plan-type (totale herhaling). Binnen de andere conditie wordt onderscheid gemaakt tussen geen feedback en kennis van correcte resultaten. In totaal derhalve vier groepen.

De meting der modelvariabelen uit het model van Bloom geschiedt als volgt:

1. het cognitief begingedrag (CBG): gemeten door middel van een vreemde-talen-aptitude-test en een voortoets, waarin alle cursusdoelen getoetst worden;
2. de affectieve leerlingkenmerken: gemeten door middel van een aangepaste versie van een attitudeschaal;
3. de kwaliteit van de instructie: wel/geen feedback en correctie;
4. de cognitieve leerresultaten: gemeten door middel van een summatieve na- en retentietoets en twee summatieve tussen-toetsen;
5. de affectieve leerresultaten: gemeten door middel van een tweede afname van de attitudeschaal;
6. de leertijd: gemeten door bepaling van de totaal-tijd besteed aan het bestuderen van de leertaak, maken van formatieve toetsen, en het uitvoeren van correctieve procedures.

De resultaten van het experiment kunnen als volgt samengevat worden:

- het model zoals in dit onderzoek gerealiseerd, gedraagt zich zoals in theorie is voorspeld;
- het CBG verklaart in de beheersingslerenconditie 46.6% en in de andere conditie 62.2% van de variantie op een leerdoelgebonden natoets;
- na correctie voor leertijdverschillen zijn deze percentages resp. 39.3% en 67.8%: de invloed van voorafgaand leren wordt in de beheersingslerenconditie dus beperkt;

- er kon geen interactie aangetoond worden tussen het CBG en de correctieve procedures;
- de affectieve leerlingkenmerken verklaren geen variantie op de natoets, ook niet na correctie voor leertijd;
- de variabele kwaliteit van de instructie verklaart 8.2% van de variantie, na correctie voor leertijd wordt dit gereduceerd tot 1%.

Geconcludeerd wordt dat het model wellicht bijstelling behoeft betreffende de rol van de affectieve leerlingkenmerken, terwijl ook de percentages verklaarde variantie aanpassing behoeven. Voorts is het zinvol in het model consequent rekening te houden met de leertijd. De aangebrachte variaties in feedback leverden vrijwel geen significante resultaten op. In de bespreking van de onderzoeksresultaten worden enige kanttekeningen geplaatst bij de opzet van het experiment, terwijl ook de mogelijkheid wordt gesuggereerd van interactie tussen type feedback en correctie enerzijds en cognitief begingedrag anderzijds. Deze suggestie vormt mede de aanleiding tot het tweede experiment.

Hoofdstuk 8 beschrijft de opzet en uitvoering van een experiment waarin wordt nagegaan of er een relatie bestaat tussen het niveau van het CBG en de leerprestaties binnen een beheersingslerenconditie. Voorts wordt nagegaan of er sprake is van interactie tussen CBG, moeilijkheidsgraad van de leer-taak en type correctieve procedure.

De opzet van het tweede experiment is in grote lijnen gelijk aan die van het eerste. Op basis van de meting van het CBG zijn de proefpersonen verdeeld in hoog- en laag-CBG, en vervolgens aan de diverse condities toegewezen. Zowel voor de hoog-, als voor de laag-CBG-proefpersonen zijn drie condities onderscheiden. De eerste conditie krijgt totale herhaling (Keller-plan-type) als correctivum, de tweede conditie partiële herhaling (Bloom-type), terwijl de derde conditie totale herhaling wordt aangeboden bij de moeilijke en partiële herhaling bij de eenvoudige leertaken. Op basis van de resultaten van het eerste experiment zijn twee moeilijke en twee eenvoudige leertaken geselecteerd. Het onderwijsleer-proces voor alle condities is opgezet conform de principes van het beheersingsleren. De affectieve leerlingkenmerken zijn niet gemeten. De meting van de cognitieve leerresultaten geschiedt wederom door middel van een summatieve na- en retentietoets. Ook de totale leertijd is weer bepaald. De resultaten van het tweede experiment kunnen als volgt worden samengevat:

- zowel op de na- als op de retentietoets is er sprake van een significant CBG-hoofdeffect: de beheersingsleren-strategie slaagt er niet in het effect van voorafgaand

- leren volledig terug te dringen;
- geen significante effecten kunnen aangetoond worden voor de condities, noch voor de interactie tussen condities en CBG;
- het onderscheid tussen totale en partiële herhaling blijkt geen invloed te hebben op de summatieve toets-scores;
- het afhankelijk van de moeilijkheidsgraad van de leerstof nu eens totale, dan weer partiële herhaling aanbieden blijkt zowel voor leerlingen met een hoog-, als voor leerlingen met een laag-CBG lagere resultaten op te leveren dan voor elk correctivum apart;
- na correctie voor leertijdverschillen blijkt dat partiële herhaling tot een efficiënter leerproces leidt dan totale herhaling, zonder dat de leerprestaties overigens significant beter zijn.

Hoofdstuk 9 tenslotte bevat een algemene discussie waarin de onderzoeksresultaten worden samengevat, en waarbij tevens de beheersingslerenstrategie in een ruimer theoretisch kader wordt geplaatst. Dit kader wordt gevormd door theorieën en modellen betreffende de zogenaamde 'direct instruction'-benadering. Ook wordt ingegaan op de relatie tussen beheersingsleren en aptitude-treatment-interaction. Een bespreking van enkele praktische implicaties van het uitgevoerde onderzoek besluit dit hoofdstuk.

SUMMARY

The purpose of this thesis is to test empirically a model dealing with learning at school introduced by Bloom (1976). Attention is paid especially to the variance in pupil achievement on a summative test which is explained by the different model variables. The teaching-learning process is structurized following the principles of the mastery learning strategy.

Broadly speaking the thesis consists of three sections.

The first section consisting of chapters 1 to 4 is a theoretical section in which Bloom's model is explained and attention is paid to individual differences between pupils. This section also includes the research problems of this study.

The second section, chapters 5 and 6, describes the course used in our research and the formative tests constructed for it.

The third section, consisting of chapters 7, 8 and 9 is the research section in which two experiments are described.

Chapter 1 presents a short account of the relationship between instructional psychology and educational science. Furthermore we discuss the development of miniature theories regarding aspects of teaching-learning situations.

Chapter 2 contains an exposition of individual differences between pupils and of educational adaptive models. We show how teaching by means of variation in learning goals, forms of instruction or both can be adapted to individual differences between pupils.

This chapter also pays attention to Carroll's model of school learning (1963), where differences between pupils are defined in terms of time. The differences in learning achievement between pupils depend not only on the characteristics of the individual pupils but also on the characteristics of the instructional process, i.e. the quality of the instruction. The *strategy of mastery learning* (Bloom, 1968) inspired by Carroll's model, reflects an opinion with regard to the optimal shaping of the teaching-learning-process.

The strategy is characterized by at least five aspects, i.e.:

1. the objectives of a course are fixed;
2. the subject matter is divided into learning tasks; learning task k must be mastered before learning task $(k + 1)$ is started;
3. after each learning task, a test is given and feedback presented;
4. in the case of inadequate mastery of the learning task a corrective is presented;

5. there is temporal differentiation, but no level-differentiation.

The mastery learning strategy takes individual differences into account by applying temporal differentiation, while in the case of pupils who have not yet mastered the subject matter after the (uniform) instructional process, various corrective measures can be taken. Finally the Keller-plan - a form of mastery learning - is discussed in this chapter and some empirical research with regard to the strategy of mastery learning is gone into.

In *Chapter 3* a survey of Bloom's (1976) model is presented. The model is meant to be an explanatory model explaining differences in pupil's learning results. Pupil characteristics are to be found in the input-side, i.e. cognitive entry behaviour and the affective pupil characteristics. Cognitive entry behaviour can account for approximately 50% of the variance in pupil achievement on the summative test. Affective pupil characteristics refer to the attitude and the interest of the pupils in beginning special learning tasks. These characteristics can account for approximately 25% of the variance in learning achievement on a summative test. Quality of instruction is taken as being the most important variable in the process section of the model.

A qualitatively good instructional process is one which satisfies the principles of mastery learning. This variable can account for 20-25% of the variance in pupil achievement on the summative test. The complete model accounts for around 90% of the variance in summative test achievement.

The output-side of the model is concerned with cognitive learning achievement, affective learning results and learning time. Moreover in this chapter separate attention is paid to different sections of the model, for example to feedback, to corrective procedures and to learning time. At the same time some critical question-marks are placed besides the percentages of variance explained by the model.

Chapter 4 concludes the first section of the thesis in which the formulations of the research problem are worked out in more detail. The outset of the formulations of the problems is based on the question to what extent in a realistic situation, variables from Bloom's model are able to account for variance in pupil achievement. At the same time research questions are formulated with regard to the supposed effects of mastery learning strategy. For example to what extent does the application of this strategy lead to a decrease in the relationship between aptitude and learning achievement? Moreover, learning time which by the application of feedback and corrective procedures can form an alternative explanation

for the possible differences in learning achievement, is taken into account.

The second section of the thesis describes the course and the formative tests used in the study.

Chapter 5 describes the French course designed for senior students who possess some previous knowledge of the subject. The course consists of seven lessons on grammatical problems. The whole course can be done individually and a pupil can work at his own pace. Each lesson ends with a formative test.

Chapter 6 contains the psychometric analysis of the formative tests. The tests can be regarded as being criterion referenced (content specific) tests. The chapter consists of two sections. The first section deals with modern views and procedures with regard to determining cut-off scores as well as to the estimation of the consistency of criterion-referenced tests. The second section of this chapter deals with the concrete application of these techniques in the tests which we used. We used three methods to define the cut-off scores:

1. the intuitive 80% method: the pupils being tested had to score at least 80% of the items correct (this is the method used in the experiment);
2. the Bayes-procedure: that cut-off score was chosen whereby the average utility of all decisions made is as large as possible;
3. the maximin-procedure: that cut-off score is chosen whereby the utility of the decision made with the smallest utility of all, is still as great as possible.

For the cut-off scores defined on the basis of the first method, the Bayes- and maximin-utility is also calculated in order to compare them. Generally speaking our research shows that the utilities of the various cut-off scores on the formative tests seem to be sufficiently high.

Two methods were used to estimate consistency:

1. Huynh's method (1976) whereby coefficient κ (Kappa) is calculated.
2. Subkoviak's method (1976) whereby coefficient p_0 is calculated.

Consistency estimation is done on all three of the cut-off scores defined. In general the consistency-coefficients appear to be of adequate size.

Finally, the third section of the thesis describes the empirical research which we carried out. The purpose of the experiment described in *chapter 7* is the systematic testing of Bloom's model. At the same time attention is paid to some different forms of feedback and corrective procedures.

Two conditions are distinguished: one whereby instruction is given structured according to mastery learning principles and another condition where this is not the case. The distinguishing element between both conditions is the presence or absence of feedback and correctives. In the mastery learning condition two types of correctives can be distinguished: a Bloom-type (partial rehearsal with the use of a diagram) and a Keller-plan-type (complete rehearsal). Within the other condition, distinction is made between no feedback and knowledge of correct results. Consequently there are four groups.

The measuring of model variables from Bloom's model is done as follows:

1. the cognitive entry behaviour: measured with a foreign language-aptitude test and a pretest in which all aspects of the course are tested;
2. the affective pupil characteristics measured with an adapted version of an attitude scale;
3. the quality of the instruction: with or without feedback and correction;
4. the cognitive learning results: measured with the use of a summative post- and retention test and two summative intermediate tests;
5. the affective learning results: measured by a second taking of the attitude-scale;
6. the learning time: measured by determining the total time, spent on studying the learning task, the time spent in doing the formative tests and the time spent in carrying out the corrective procedures.

The results of the experiment can be summarized as follows:

- The model as it appears in this research performs in the same way as we had theoretically expected.
- The cognitive entry behaviour (CEB) explains 46.6% in the mastery learning condition and in the other condition 62.2% of the variance on a criterion-referenced post-test.
- Correction having been made for learning time differences these percentages are then 39.3% and 67.8% respectively; thus influence of previous learning is limited in the mastery learning condition.
- No interaction is indicated between CEB and the corrective procedures.
- The affective pupil characteristics explained no variance in the post-test, not even after learning-time correction.
- The variable quality of instruction accounts for 8.2% of the variance; after correction for learning-time this is reduced to 1%.

Thus it can be concluded that perhaps the model needs

modification with regard to the role of the affective pupil characteristics while the percentages variance accounted for also need to be modified. Furthermore it is desirable to consistently take learning time into account in the model. Variation introduced in feedback showed almost no significant results. In the discussion about the results of the research some critical remarks about the design of the experiment are made, while it is suggested that there is also a possible interaction, on one side between feedback type and correction, and on cognitive entry behaviour on the other. This suggestion is responsible for the experiment described in the following chapter.

Chapter 8 describes the design and carrying out of an experiment in which we checked whether a relationship exists between the level of the CEB and learning achievement within a mastery learning condition. Further we checked as to whether there is a question of interaction between CEB, difficulty level of the learning task and corrective procedure type.

The design of the second experiment is to a large extent similar to the design of the first one. On the basis of the CEB measurement the subjects are divided into high and low CEB and then assigned to the various conditions. In the case of both high and low CEB subjects three conditions are distinguished. The first condition receives total rehearsal (Keller-plan-type) as corrective, the second condition gets partial rehearsal (Bloom-type) while in the case of the third condition total rehearsal is offered in the case of difficult learning tasks and partial rehearsal in the case of easy learning tasks. On the basis of the results of the first experiment two difficult and two easy learning tasks are selected. The teaching-learning-process for all conditions is designed in accordance with mastery learning principles. The affective pupil characteristics have not been measured. Once again a summative post- and retention test is used to measure the cognitive learning results. Again total learning time is determined. The results of the second experiment can be summarized as follows:

- In both the post- and retention test a significant CEB main-effect is indicated: mastery learning strategy does not succeed in completely suppressing the effect of previous learning.
- No significant effects are indicated for the conditions, or for the interaction between the conditions and CEB.
- The difference between total and partial rehearsal appears to have no influence on the summative test-scores.
- Depending on the level of difficulty of the learning tasks, sometimes offering total and sometimes offering partial

rehearsal, mastery learning strategy appears to give lower results both for pupils with a high and for pupils with a low CEB than for each corrective seperately.

- After correction for learning time differences it appears that partial rehearsal leads to a more efficient learning *process* than total rehearsal, without learning *achievement* being significantly better.

Finally, *chapter 9* consists of a general discussion in which the research results are summarized and whereby mastery learning strategy is placed in a wider theoretical framework. This framework consists of theories and models concerning the so called 'direct instruction' approach. Also, the relationship between mastery learning and aptitude-treatment-interaction is mentioned. This chapter ends with a discussion about some of the practical implications of the research carried out.

LITERATUUR

- Anderson, L.W.,
An empirical investigation of individual differences in time to learn. *Journal of Educational Psychology*, 1976, 68, 226-233.
- Anderson, L.W.,
New directions for research on instruction and time-on-task. Boston: AERA-paper, 1980.
- Anderson, R.C. & Faust, G.W.,
The effects of strong formal prompts in programmed instruction. *American Educational Research Journal*, 1967, 4, 345-352.
- Anderson, R.C., Kulhavy, R.W. & Andre, T.,
Feedback procedures in programmed instruction. *Journal of Educational Research*, 1971, 62, 148-156.
- Anderson, R.C., Kulhavy, R.W. & Andre, T.,
Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology*, 1972, 63, 186-188.
- Arlin, M.N.,
Learning rate and learning rate variance under mastery learning conditions. Univ. of Chicago: Unpubl. diss., 1973.
- Ausubel, D.P.,
Educational psychology: A cognitive view. New York: Holt, Rinehart and Winston, 1968.
- Bennett, N.,
Teaching styles and pupil progress. London: Open Books, 1976.
- Berk, R.A.,
Determination of optimal cutting scores in criterion-referenced measurement. *Journal of Experimental Education*, 1976, 45, 4-9.
- Binor, S.,
The relative effectiveness of mastery learning strategies in second language acquisition. Univ. of Chicago: Unpubl. master's dissertation, 1974.

- Birnbaum, A.,
Some latent trait-models and their use in inferring an examinee's ability. In: Lord, F.M. & Novick, M.R., *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.
- Block, J.H.,
The effects of various levels of performance on selected cognitive, affective and time variables. Univ. of Chicago: Unpubl. diss. 1970.
- Block, J.H. (Ed.),
Mastery learning: Theory and practice. New York: Holt, Rinehart & Winston, 1971.
- Block, J.H.,
Student learning and the setting of mastery performance standards, *Educational Horizons*, 1972, 50, 183-191.
- Block, J.H. (Ed.),
Schools, society and mastery learning. New York: Holt, Rinehart & Winston, 1974.
- Block, J.H.,
Success, rate. In: Denham, C. & Liberman, A. (Eds.), *Time to learn*. Washington, D.C.: The National Institute of Education, 1980, 95-106.
- Block, J.H. & Anderson L.W.,
Mastery learning in classroom instruction. New York: Macmillan Publishing Co., Inc., 1975.
- Block, J.H. & Burns, R.B.,
Time in school learning: An instructional psychologist's perspective. Washington, D.C.: AERA-paper, 1975.
- Block, J.H. & Burns, R.B.,
Mastery learning. In: L.S. Shulman (Ed.), *Review of research in education*, 4, Itasca, Ill.: Peacock Publ., 1977, 3-49.
- Bloom, B.S. et al.,
Taxonomy of educational objectives-the classification of educational goals, Handbook I, Cognitive domain. New York: Mc. Kay Company, 1956.

- Bloom, B.S.,
Learning for mastery. U.C.L.A.-C.S.E.I.P.-*Educational comment*, 1968, Vol. 1, no. 2.
- Bloom, B.S.,
Recent developments in mastery learning. *Educational Psychologist*, 1973, 10, 204-221.
- Bloom, B.S.,
Time and learning. *American Psychologist*, 1974, Vol. 29, 682-688.
- Bloom, B.S.,
Human characteristics and school learning. New York: McGraw-Hill, 1976.
- Bloom, B.S.,
The new direction in educational research: alterable variables. *The Kappan*, 1980 (febr.).
- Bloom, B.S., Hastings, J. Th. & Madaus, G.F.,
Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.
- Borich, G.D. & Godbout, R.C.,
Extreme groups designs and the calculation of statistical power. *Educational and Psychological Measurement*, 1974, 34, 663-675.
- Boyd, W.M.,
Repeating questions in prose learning. *Journal of Educational Psychology*, 1973, 64, 31-38.
- Bruner, J.S.,
Toward a theory of instruction. Cambridge Mass.: Harvard University Press, 1966.
- Braak, L.H.,
Geïndividualiseerde onderwijssystemen: constructie en besturing. Eindhoven: Techn. Hogeschool, 1974 (Proefschrift).
- Bruyne, H.C.D. de,
Blokken in het onderwijs. Verkenningen op het terrein van beheersingsleren. Groningen: H.D. Tjeenk Willink, 1976.

- Burrows, C.K. & Okey, J.R.,
The effects of a mastery learning strategy on achievement. Washington D.C.: AERA-paper, 1975
- Buss, A.R. & Poley, W.,
Individual differences: traits and factors. New York: Gardner Press, Inc., 1976.
- Campbell, D.T. & Stanley, J.C.,
 Experimental and quasi-experimental designs for research on teaching. In: Gage, N.L. (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.
- Carlson, J. & Minke, K.,
 Fixed and ascending criteria for unit mastery learning. *Journal of Educational Psychology*, 1975, 67, 96-101.
- Carroll, J.B.,
 The prediction of success in intensive language training. In: Glaser, R. (Ed.), *Training research and education.* Pittsburg: University of Pittsburgh Press, 1962.
- Carroll, J.B.,
 A model of school learning. *Teachers' College Record*, 64, 1963, 723-733.
- Carroll, J.B.,
The teaching of French as a foreign language in seven countries: International Studies in Evaluation (IEA) V. New York: John Wiley & Sons, Inc., 1975.
- Carroll, J.B. & Sapon, S.M.,
Modern language aptitude test (MLAT). New York: The Psychological Corporation, 1959.
- Carver, R.P.,
 Special problems in measuring change with psychometric devices. In: *Evaluative research: Strategies and methods.* Pittsburgh, Pa.: American Institutes of Research, 1970, 48-63.
- Cobb, J.A.,
 Relationship of discrete classroom behaviors to fourth grade academic achievement. *Journal of Educational Psychology*, 1972, 63, 74-80.

- Cohen, J.,
A coefficient of agreement for nominal scales.
Educational and Psychological Measurement, 1960,
20, 37-46.
- Cooley, W.W. & Lohnes, P.R.,
Evaluation research in education. New York: Irvington
Publ., Inc., 1976.
- Cooley, W.W. & Mao, B.J.,
The sample of classroom-time observed. *The Journal of
Classroom Interaction*, 1981, 17, 31-36.
- Corte, E. de, Geerligs, C.T., Lagerwey, N.A.J., Peters, J.J. &
Vandenberghe, R.,
Beknopte didaxologie. Groningen: Wolters-Noordhoff,
1981 (5e druk).
- Crombag, H.F.M.,
ATI: Perhaps not such a good idea after all. *Tijdschrift
voor Onderwijsresearch*, 1979, 4, 176-183.
- Cronbach, L.J.,
The two disciplines of scientific psychology. *American
Psychologist*, 1957, 12, 671-684.
- Cronbach, L.J.,
How can instruction be adapted to individual differences?
In: Gagné, R.M. (Ed.), *Learning and individual
differences*, Columbus (Ohio): Merrill, 1967, 23-39.
- Cronbach, L.J. & Furby, L.,
How we should measure change - or should we?
Psychological Bulletin, 1970, 74, 68-80.
- . Cronbach, L.J. & Snow, R.E.,
Aptitudes and instructional methods. New York:
Irvington Publishers, Inc., 1977.
- Denham, C. & Lieberman, A.,
Time to learn. Washington, D.C.: The National
Institute of Education, 1980.
- Dixon, W.J. & Brown, M.B.,
BMDP-79: Biomedical Computer Programs, P-series.
Berkeley's University of California Press, 1979.

- Drenth, P.J.,
Inleiding in de testtheorie. Deventer: Van Loghum
 Slaterus, 1975.
- Drenth, P.J.D. & Wieringen, P.C.W. van,
VAT'69: Verbale aanlegtestserie 1969. Amsterdam:
 Swets & Zeitlinger, 1969.
- Elashoff, J.D.,
 Analysis of covariance: A delicate instrument. *American Educational Research Journal*, 1969, 6, 383-401.
- Faw, H.W. & Waller, I.G.,
 Mathemagenic behaviours and efficiency in learning from
 prose. *Review of Educational Research*, 1976, 46,
 691-720.
- Fisher, C.W. et al.,
 Teaching behaviors, academic learning time, and student
 achievement: an overview. *The Journal of Classroom Interaction*, 1981, 17, 2-5.
- Frase, L.T.,
 Effects of question location, pacing and mode upon
 retention of prose material. *Journal of Educational Psychology*, 1968, 59, 244-249.
- Gagné, R.M.,
The conditions of learning (2nd ed.), New York:
 Holt, Rinehart & Winston, 1970.
- Gagné, R.M.,
Essentials of learning for instruction, Hinsdale,
 Ill.: Dryden Press, 1974.
- Gagné, R.M. & Briggs, L.J.,
Principles of instructional design. New York: Holt,
 Rinehart & Winston, 1974.
- Gilman, D.A.,
 Comparison of several feedback methods for correcting
 errors by computer-assisted instruction. *Journal of Educational Psychology*, 1969, 60, 503-508.
- Glaser, R.,
 Instructional technology and the measurement of learning
 outcomes: some questions. *American Psychologist*,

1963, 18, 519-521.

Glaser, R.,

Individuals and learning: The new aptitudes. *Educational Researcher*, 1972, 1, 5-13.

Glaser, R.,

Educational psychology and education, *American Psychologist*, 1973, 28, 557-566.

Glassnapp, D.R., Deaton, W.L. & Sokolove, H.,

Aptitude, achievement and time to criterion. Toronto: AERA-paper, 1978.

Glassnapp, D.R., Poggio, J.P. & Ory, J.C.,

Cognitive and affective consequences of mastery and non-mastery instructional strategies. Washington, D.C.: AERA-paper, 1975.

Glass, G.V.,

Primary, secondary and meta-analysis, *Educational Researcher*, 5, 1976.

Glass, G.V.,

Integrating findings: The meta-analysis of research. *Review in Education*, 1977, 5, 351-379.

Good, T.L.,

Teacher effectiveness in the elementary school. *Journal of Teacher Education*, 1979, 30, 52-64.

Gump, P.V.,

The classroom behavior setting: Its nature and relation to student behavior. Lawrence (Kansas): Dept. of Psychology, University of Kansas, 1979 (ERIC ED 015-515).

Hambleton, R.K., Swaminathan, H., Algina, J. & Coulson, D.B.,

Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research*, 1978, 48, 1-47.

Hartley, J. & Davies, I.K.,

Preinstructional strategies: The role of pretests, behavioral objects, overviews and advance organizers. *Review of Educational Research*, 1976, 46, 239-265.

- Hays, W.L.,
Statistics. London: Holt, Rinehart & Winston, Inc.
 1969.
- Hilgard, E.R. & Bower, G.H.,
Theories of learning. Englewood Cliffs: Prentice
 Hall, Inc., 1975.
- Hunt, D.E. & Sullivan, E.V.,
Between psychology and education. Hinsdale, Ill.:
 Dryden, 1974.
- Huynh, H.,
 On the reliability of decisions in domain-referenced
 testing. *Journal of Educational Measurement*, 1976,
 13, 253-264.
- Huynh, H.,
 Reliability of multiple classifications.
Psychometrika, 1978, 43, 317-325.
- Huynh, H.,
 A nonrandomized minimax solution for passing scores in the
 binomial error model. *Psychometrika*, 1980, 45,
 167-182.
- Idenburg, Ph. J.,
Theorie van het onderwijsbeleid, Groningen: Wolters-
 Noordhoff, 1971.
- Jakobovits, L.A.,
*Foreign language learning. A psycholinguistic analysis
 of the issues*, Rowley, 1970.
- Keller, F.S.,
 Goodbye, teacher ... *Journal of Applied Behavioral
 Analysis*, 1968, 1, 79-89.
- Keller, F.S.,
 Ten years of personalized instruction. *Teaching of
 Psychology*, 1974, 1, 4-9.
- Keller, F.S. & Sherman, J.G.,
The Keller Plan handbook. Menlo Park (Cal.): W.A.
 Benjamin, Inc., 1974.

- Kerlinger, F.N. & Pedhazur, E.J.,
Multiple regression in behavioral research. New York:
 Holt, Rinehart & Winston, 1973.
- Kirschner, P.A.,
*Tekstkenmerken en leerprocessen. Deel II: Toegevoegde
 vragen*. Groningen: Wolters-Noordhoff (afd. Studie &
 Onderzoek), 1980.
- Klerk, L.F.W. de,
Van leerpsychologie naar onderwijspsychologie.
 Deventer: Van Loghum Slaterus, 1974 (Openbare rede).
- Klerk, L.F.W.,
 De invloed van feedback op het onderwijsleerproces.
Pedagogische Studiën, 1977, 54, 8-13.
- Klerk, L.F.W. de,
Inleiding in de onderwijspsychologie. Deventer: Van
 Loghum Slaterus, 1979a.
- Klerk, L.F.W. de,
 ATI, perhaps a good idea after all! *Tijdschrift voor
 Onderwijsresearch*, 1979b, 4, 191-194.
- Klerk, L.F.W. de,
 Psycho-educational research: a need for miniature theories.
 In: Boekaerts, M. & L.F.W. de Klerk (eds.),
*Psycho-educational research. Communication and
 cognition*. Gent, 1982.
- Knibbeler, W.,
Kwantitatieve analyse van een cursus Frans.
 Collection d' 'Etudes linguistiques' 18, Brussel: AIMAV,
 1975.
- Knibbeler, W.
*Het onafhankelijk leren van een vreemde taal door
 volwassenen*. Groningen: H.D. Tjeenk Willink bv., 1976.
- Knippenberg, A. van & Siero, F.,
Multivariate analyse. Deventer: Van Loghum Slaterus,
 1980.
- Kounin, J.S.,
Discipline and group management in classrooms. New
 York: Holt, Rinehart & Winston, 1970.

- Kulhavy, R.W.,
Feedback in written instruction. *Review of Educational Research*, 1977, 47, 211-232.
- Kulhavy, R.W. & Anderson, R.C.,
Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, 1972, 63, 505-512.
- Kulik, J.A., Kulik, C.L. & Carmichael, K.,
The Keller-plan in science teaching. *Science Magazine*, 1974, 379-383.
- Linden, W.J. van der,
Psychometric contributions to the analysis of criterion-referenced measurements. Amsterdam, 1980 (Proefschrift).
- Linden, W.J. van der,
Boekbespreking van: Berk, R.A. (ed.), Criterion-referenced measurement: The state of the art, Baltimore: The John Hopkins Univ. Press, 1980. In: *Tijdschrift voor Onderwijsresearch*, 1981, 6, 299-301.
- Linn, R.L. & Slinde, J.A.,
The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research*, 1977, 47, 121-150.
- Lodewijks, J.G.L.C.,
Leerstofsequenties: van conceptueel netwerk naar cognitieve structuur. Tilburg, 1981 (Proefschrift).
- Lord, F.M. & Novick, M.R.,
Statistical theories of mental test scores. Reading, Mass.: Addison Wesley Publishing Company, 1968.
- McDonald, F.J. & Allen, D.,
An investigation of presentation, response and correction factors in programmed instruction. *Journal of Educational Research*, 1962, 55, 502-507.
- Merrill, M.D.,
Specific review versus repeated presentation in a programmed imaginary science. *Journal of Educational Psychology*, 1970, 61, 392-399.

- Merrill, M.D., Barton, K. & Wood, L.E.,
Specific review in learning a hierarchical imaginary science. *Journal of Educational Psychology*, 1970, 61, 102-109.
- Meskauskas, J.A.,
Evaluation-models for criterion-referenced testing: views regarding mastery and standard-setting. *Review of Educational Research*, 1976, 46, 133-158.
- Millman, J.,
Criterion-referenced measurement. In: Popham, W.J. (Ed.), *Evaluation in education: Current applications*. Berkely, Calif.: McCutchan, 1974.
- Möller, C.,
Technik der Lernplanung. Weinheim und Basel: Verslag Julius Beltz, 1973 (4e druk).
- Mood, A.M., Graybill, F.A. & Boes, D.C.,
Introduction to the theory of statistics, Tokyo: McGraw-Hill, Kogakusha, 1974.
- Nuy, M.J.G.,
Interne differentiatie: over het ontwerpen van geïndividualiseerde onderwijsarrangementen. 's-Hertogenbosch: Katholiek Pedagogisch Centrum, 1981 (Proefschrift).
- Nuy, M.J.G. & Timmer, J.,
Enkele ervaringen bij de introductie van mastery learning in het voortgezet onderwijs. In: Warries, E. (e.a.), *Beheersingsleren een leerstrategie*. Groningen: Wolters-Noordhoff, 1979, 95-122.
- Nijhof, W.J.,
Interne differentiatie als een innovatie. 's-Gravenhage: Staatsuitgeverij, 1978 (SVO-Reeks nr. 9).
- Okey, J.R., Brown, J.L. & Fiel, R.L.,
Diagnostic evaluation methods in individualized instruction. *Science Education*, 1972, 56, 207-212.
- Parreren, C.F. van,
Onderwijsproceskunde. Groningen: Wolters-Noordhoff, 1981.

- Parreren, C.F. & Carpay, J.A.M.,
Sovjet-psychologen aan het woord. Groningen: Wolters-Noordhoff, 1972.
- Parreren, C.F. & Carpay, J.A.M.,
Sovjet-psychologen over onderwijs en cognitieve ontwikkeling. Groningen: Wolters-Noordhoff, 1980.
- Peeck, J.,
 Effects of prequestions on delayed retention of prose material. *Journal of Educational Psychology*, 1970, 61, 241-246.
- Peeck, J.,
 Preinstructional strategies and extra reading time in learning from text. *Tijdschrift voor Onderwijsresearch*, 1977, 2, 202-207.
- Peterson, P.L.,
 Book review: Bloom, B.S., Human characteristics and school learning. New York: McGraw-Hill, 1976. In: *American Educational Research Journal*, 1977, 14, 73-75.
- Plomp, Tj.,
De ontwikkeling van een individueel studie-systeem. Groningen: H.D. Tjeenk-Willink bv., 1974.
- Plomp, Tj. & Pilot, A.,
 Beheersingsleren in het tertiair onderwijs. In: Warries, E. (e.a.), *Beheersingsleren een leerstrategie*. Groningen: Wolters-Noordhoff, 1979, 123-139.
- Poortvliet, G.A.,
 Mastery learning, resultaten van buitenlands onderzoek. In: Warries, E. (e.a.), *Beheersingsleren een leerstrategie*. Groningen: Wolters-Noordhoff, 1979, 31-41.
- Rapport:
Model van een opleiding onderwijskunde. 's-Gravenhage: Academische Raad, 1973.
- Rice, M.J.,
Variables in mastery learning. San Francisco: AERA-paper, 1976.

- Rookhuijzen, R.F. van, Plomp, Tj. & Pilot, A.,
Individuele studiesystemen in het tertiair onderwijs. Een overzicht. Groningen: Wolters-Noordhoff, 1976 (OTO-cahier No. 3).
- Rosenberg, M.J.,
 Cognitive structure and attitudinal affect. In: Fishbein, M. (ed.), *Readings in attitude theory and measurement*, New York: 1967, 325-331.
- Rosenshine, B.V.,
 Content, time and direct instruction. In: Peterson, P.L. & Walberg, H.J. (Eds.), *Research on teaching*, Berkeley (Cal.): McCutchan Publ. Corp., 1979, 28-56.
- Rosenshine, B.V.,
 How time is spent in elementary classrooms. *The Journal of Classroom Interaction*, 1981, 17, 16-25.
- Rosenshine, B. & Furst, N.F.,
 The use of direct observation to study teaching. In: Travers, R.M.W. (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally, 1973, 122-183.
- Rothkopf, E.Z.,
 Learning from written material: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal*, 1966, 3, 241-249.
- Rothkopf, E.Z. & Bisbicos, E.E.,
 Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, 1967, 58, 56-61.
- Ryan, B.A.,
PSI: Keller's personalized system of instruction: An appraisal. Washington, D.C.: American Psychological Association, 1974.
- Sassenrath, J.M. & Younge, G.D.,
 Effects of delayed information feedback, feedback cues, retention set, and delayed retention. *Journal of Educational Psychology*, 1969, 60, 174-177.
- Schmid-Schönbein, G.,
 Single-case experimental designs in remedial teaching of

- English as a foreign language. *Instructional Science*, 1980, 9, 183-194.
- Semb, G.,
The effects of mastery criteria and assignment length on college students test performance. *Journal of Applied Behavior Analysis*, 1974, 7, 61-69.
- Simons, P.R.J.,
Vergelijkenderwijs: onderzoek naar de invloed van metaforen op het leren. Tilburg, 1980 (Proefschrift).
- Snow, R.E.,
Learning and individual differences. In: Shulman, L.S. (Ed.), *Review of research in education* (Vol. 4), Itasca (Ill.): F.C. Peacock & Co, 1977a.
- Snow, R.E.,
Individual differences and instructional theory. *Educational Researcher*, 1977b, 6, 11-15.
- Span, P.,
De ontwikkeling van een onderwijsleerpsychologie. *Pedagogische Studiën*, 1982, 59, 197-199.
- Stallings, J.A.,
Implementation and child effects of teaching practices in Follow through classrooms. *Monographs of the Society for Research in Child Development*, 1975, 40, 1-119.
- Sturges, P.T.,
Verbal retention as a function of the information and delay of informative feedback. *Journal of Educational Psychology*, 1969, 60, 11-14.
- Subkoviak, M.J.,
Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement*, 1976, 13, 265-276.
- Subkoviak, M.J.,
Empirical investigation of procedures for estimating reliability for mastery tests. *Journal of Educational Measurement*, 1978, 15, 111-116.
- Subkoviak, M.J.,
Decision-consistency approaches. In: Berk, R.A. (ed.),

Criterion-referenced measurement: The state of the art. Baltimore & London: The John Hopkins Univ. Press, 1980, 129-185.

Sullivan, H.J., Schutz, R.E. & Baker, R.L.
Effects of systematic variations in reinforcement contingencies on learner performance. *American Educational Research Journal*, 1971, 8, 135-142.

Swanson, D.H.,
A comparison of mastery learning feedback systems affecting achievement in chemistry. New York: AERA-paper, 1977.

Tait, K., Hartley, J.R. & Anderson, R.C.,
Feedback procedures in computer-assisted arithmetic instruction. *British Journal of Educational Psychology*, 1973, 43, 161-171.

Talyzina, N.F.,
Psychological bases of programmed instruction. *Instructional Science*, 1973, 2, 243-280.

Thorndike, R.L.,
Reading comprehension education in fifteen countries: International Studies in Evaluation (IEA). New York: John Wiley & Sons, Inc., 1973.

Tierney, M.,
A comparison of feedback/correction procedures used in mastery learning strategies. Santa Barbara: University of California, 1973 (Unpublished Master's Thesis).

Tobias, S.,
Achievement treatment interactions. *Review of Educational Research*, 1976, 46, 61-74.

Tobias, S.,
When do instructional methods make a difference? *Educational Researcher*, 1982, 11, 4-9.

Trotsenburg, E.A. van,
Ontwikkelingslijnen in het empirisch onderzoek van pedagogische en didactische vraagstukken. Amsterdam: Scholar's Press, 1972 (Proefschrift).

- Ulijn, J.,
Foutenanalyse en vreemde-taalverwerving. *Levende Talen*, 1975, 310, 50-58.
- Veenman, S.A.M.,
De 'Beginning Teacher Evaluation Study' (I) & (II), *Pedagogische Studiën*, 1980, 57, 207-217 en 273-287.
- Veldhuijzen, N.H.,
Aspecten van criterion-referenced measurement.
Specialistisch Bulletin, Nr. 1, Arnhem: Cito, z.j.
- Veldhuijzen, N.H.,
Cesuurbepaling in het beta-binomiale model..
Specialistisch Bulletin, Nr. 4, Arnhem: Cito, 1979.
- Veldhuijzen, N.H.,
Setting cutting scores: A minimum information approach.
Evaluation in Education, 1982, 5, 141-148.
- Verhoeven, A.F.M.,
Opmerkingen naar aanleiding van Crombag's artikel: 'ATI-perhaps nog such a good idea after all'. *Tijdschrift voor Onderwijsresearch*, 1979, 4, 195-198.
- Verreck, W.A.,
Individualisering in het wetenschappelijk technisch onderwijs: evaluatie van een experiment in de technische mechanica. Eindhoven: Technische Hogeschool, 1973 (Proefschrift).
- Ward, W.A.,
Introduction to linear models. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1973.
- Warries, E.,
Het relatief meten van leerprestaties in het onderwijs. *Nederlands Tijdschrift voor de Psychologie*, 1970, 25, 429-439.
- Warries, E.,
Het leermodel van Carroll en de strategie van Bloom. In: Warries, E. (e.a.), *Beheersingsleren een leerstrategie*, Groningen: Wolters-Noordhoff, 1979, 11-18.

Warries, E.

Relative measurement and the selective philosophy in education. *Evaluation in Education*, 1982, 5, 191-202.

Weeda, P.,

'Aptitude-treatment-interaction' (Verslag van een studiedag). Utrecht: Palo-memo 0186-90, 1973.

Weeda, P.,

Het differentieel effect van correctieve procedures bij de formatieve evaluatie van onderwijsleerprocessen. Utrecht: ORD-paper, 1978.

Weisgerber, R.A. (ed.),

Developments in individualised learning. Itasca, Ill.: Peacock Publishers, 1971.

Wentling, T.L.,

Mastery versus non-mastery instruction with varying test item feedback treatments. *Journal of Educational Psychology*, 1973, 65, 50-58.

Wiley, D.E. & Harnischfeger, A.,

Explosion of a myth; quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher*, 1974, 3, 7-12.

Wolters, L.,

Vragen over vragen. Nijmegen: Instituut voor Onderzoek van het Wetenschappelijk Onderwijs, 1975.

Druk: Cito huisdrukkerij

Stellingen behorend bij het proefschrift
"Beheersingsleren: het model getoetst
in de tijd".

W. C. Weeda, 29 oktober 1982

I

Het verdient aanbeveling de keuze van onderzoeksvariabelen in toegepast onderwijskundig onderzoek mede te laten afhangen van de beïnvloedbaarheid van deze variabelen in de praktijk.

Dit proefschrift

II

Onderwijsmethoden die de leerlingen aanzetten tot meer feitelijk bestede tijd ("time-on-task") hebben hogere leerprestaties tot gevolg dan onderwijsmethoden waarbij dit niet het geval is.

Dit proefschrift

III

Het is geen vanzelfsprekendheid dat psychometrische modellen van toepassing zijn op examens.

IV

Toetsen, die een integraal onderdeel vormen van een leergang en bedoeld zijn als leermiddel, dienen bij voorkeur opgebouwd te zijn uit open vragen i.p.v. uit meerkeuze items.

Tobias, S., Review of the response mode issue, Review of Educational Research, 1972, 43, 93-204

V

Leerlingen met alle items goed, of alle items fout op een studietoets zijn in het Rasch-model van onschatbare waarde.

VI

Selectieve stimulering van onderzoeksgebieden binnen het universitaire eerste-geldstroemonderzoek is slechts doelstreffend in een systeem van student-onafhankelijke financiering van universitair onderzoek.

Publikatie nr. 26 van de Raad van Advies voor het Wetenschapsbeleid, 's-Gravenhage, 1982

VII

Het verdient aanbeveling om, evenals in de Bondsrepubliek Duitsland, de vrijheid van onderzoek, naast die van woord en geschrift, in de grondwet te verankeren.

VIII

De gynaecologische polikliniek biedt goede mogelijkheden voor vroegtijdige herkenning van anorexia nervosa-patiënten.

Weeda-Mannak, W.L., Het belang van vroege herkenning van anorexia-nervosa, Tijdschrift voor psychiatrie, 1981, 23, 5-12

IX

De term "quasi" in "quasi-experimental-design" diskwalificeert onbedoeld een methode van sociaal-wetenschappelijk onderzoek.

X

Het selectief opnemen van bij een proefschrift behorende stellingen in allerlei dag- en weekbladen is een merkwaardige vorm van wetenschapsjournalistiek.

Bibliotheek K. U. Brabant



17 000 00827002 8

52137